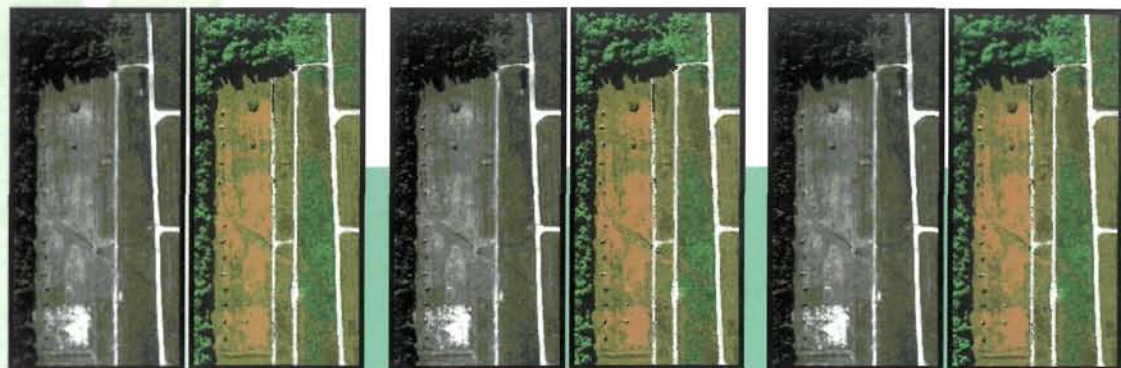
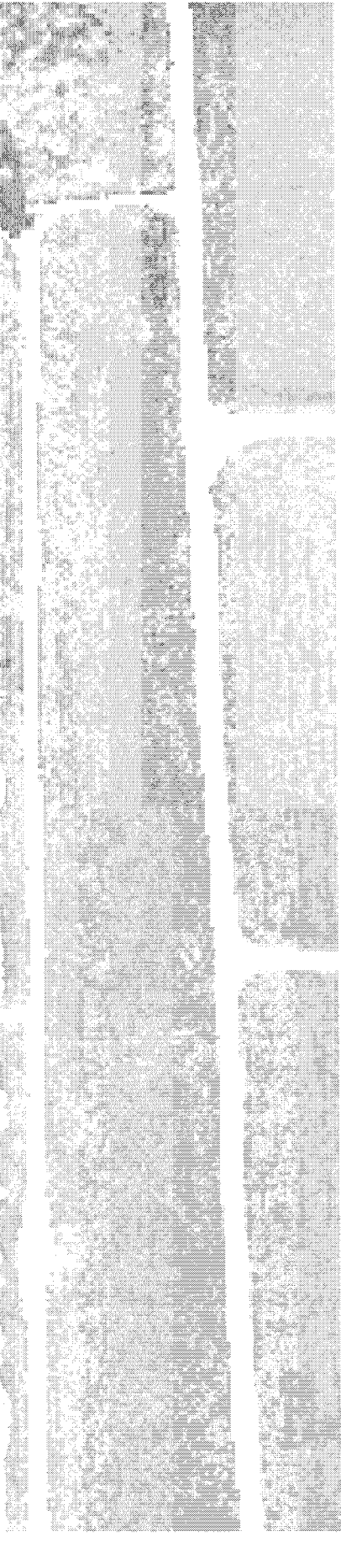


# FRONTIERS OF REMOTE SENSING INFORMATION PROCESSING



Editor: C H Chen

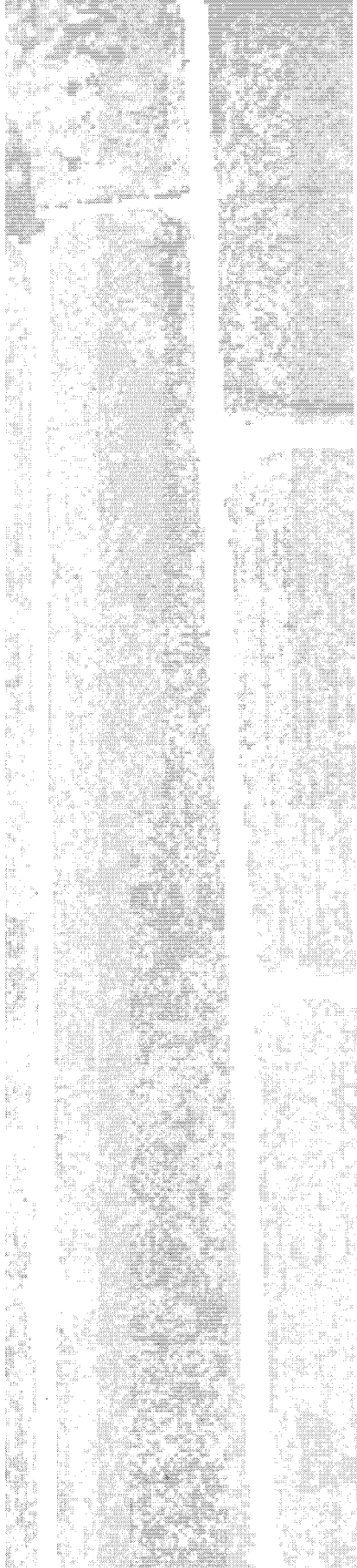
World Scientific



**FRONTIERS OF  
REMOTE SENSING  
INFORMATION  
PROCESSING**



This page is intentionally left blank



# FRONTIERS OF REMOTE SENSING INFORMATION PROCESSING

**Editor: C H Chen**

*University of Massachusetts-Dartmouth &  
Information Research Laboratory, Inc, USA*

 **World Scientific**  
*New Jersey • London • Singapore • Hong Kong*

*Published by*

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

*USA office:* Suite 202, 1060 Main Street, River Edge, NJ 07661

*UK office:* 57 Shelton Street, Covent Garden, London WC2H 9HE

**British Library Cataloguing-in-Publication Data**

A catalogue record for this book is available from the British Library.

**FRONTIERS OF REMOTE SENSING INFORMATION PROCESSING**

Copyright © 2003 by World Scientific Publishing Co. Pte. Ltd.

*All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.*

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 981-238-344-1

## PREFACE

The remote sensing area has experienced rapid growth in recent years. One major factor that impacts the growth is the numerous information processing techniques originated from the significant progress in other fields such as image and signal processing, pattern recognition, artificial intelligence, computer vision, and related areas such as neural networks, fuzzy logic, etc. There has also been a number of new mathematical approaches for example in wavelet transform, deconvolution, etc., which motivated development of more efficient information processing techniques for use in remote sensing.

The book presents chapters that highlight some frontier work in remote sensing information processing. I am very honored to have leaders in the field to prepare the chapters that present readers with these frontier developments. Although no attempt is made to cover every topic, the representative chapters in each topic should give readers a good insight of the status of remote sensing information processing. Topics are not arranged according to their order of importance. In fact all topics covered are of equal importance. Chapter lengths may vary, but they are equally significant.

The book begins with topics of general interest. With the advance of remote sensing sensors, the need for advanced information processing techniques that match the capability of the sensors is quite evident. So we begin with Chapter 1 by Prof. Richards on remote sensing sensors, on their capabilities and information processing requirements. Information processing functions such as parameter estimation and classification tasks obviously must keep up with increased complexity of sensors, and increased demand for more accurate and detailed knowledge from the data provided by the sensors. To address the issue of data fusion, Prof. Richards suggests that a more pragmatic approach is to fuse at the level of labels or decisions, and illustrates this concept by depicting the data chain model in a remote sensing system as at least one component of a knowledge system. Chapter 2 deals with the general topic of transform methods in processing remote sensing data. Proper transform methods can be useful in remote sensing for efficient data representation and reduction, noise removal, effective feature extraction for classification, and efficient image compression. For more on the transform methods, the readers are encouraged to consult the sister volume, "Information Processing for Remote Sensing", also edited by myself and published by World Scientific Publishing 1999 (ISBN 981-02-3737-5).

Since there are always a large amount of data in remote sensing, from which desired information must be extracted, data mining is a topic of increasing

interest and importance. Chapter 3 describes the work on developing a probabilistic visual grammar that supports among other things complex query scenarios that consist of many regions with different feature characteristics. The system is demonstrated with query scenarios that cannot be expressed by traditional region level approaches while the visual grammar provides accurate classification and effective retrieval. A different aspect of information mining is the use of a shape-based approach to represent geometrical patterns in agriculture and irregular boundaries of lakes, as presented in Chapter 4. The work is clearly useful also for change detection and monitoring to measure say lake change response to short or long term climatic variation.

Among the remote sensing data, SAR images have attracted the most interest because of some unique capabilities of SAR sensors. Dr. Lee and his colleagues at Naval Research Lab. presents in Chapter 5 a review of polarization orientation angle estimation and in particular a circular polarization algorithm for estimating orientation angle. For the polarimetric interferometric SAR data, Dr. Ferro-Famil and his colleagues presents in Chapter 6 the unsupervised classification of polarimetric data based on the multivariate Wishart density function of coherency matrices. A k-mean clustering algorithm using Wishart statistics segments pixels into compact clusters according to their scattering properties.

In the last ten years the wavelet transform has received enormous interest in remote sensing community. In Chapter 7, Dr. Liu and his colleagues at NASA Goddard provides a very comprehensive study of the use of Gaussian wavelet transform for near real time screening of satellite data, data reduction and image enhancement, with application to coastal monitoring (e.g. oil spills) and for ice edge and ice floe tracking, etc. Chapter 8 by Dr. Pierce and his colleagues starts with wavelet denoising overview and then presents a Markov random field (MRF) based spatially adaptive wavelet despeckling algorithm for SAR images in both intensity and amplitude formats. In Chapter 9, a redundant wavelet representation is presented using a dyadic wavelet frames. It is then extended towards orthogonal wavelet bases using the discrete wavelet transform. The representation is very suitable for image fusion and an algorithm is presented for fusion and merging of multispectral images.

Military application of remote sensing has been less reported in the literature. Chapter 10 has an excellent coverage of automated satellite interpretation algorithms and diagnostic tools that provide meteorological parameter estimation, which assist the US Navy weather observer and forecaster. The cloud classification algorithm takes advantage of all five Geostationary Operational Environmental Satellite (GOES) channels of data and makes use of the 1-nearest neighbor classification routine.

Enhancing spatial resolution of the microwave sensor data is the focus of Chapter 11. by Dr. Long who presented several algorithms for resolution



algorithms based on inverse filtering and irregular sampling reconstruction. An application example is provided on the Earth Resources Satellite Active Microwave Instrument scatterometer data.

Application of statistical pattern recognition in remote sensing dates back to the mid-60's. While the fundamental issues have been addressed and new techniques developed in statistical pattern recognition throughout the years, there has been a parallel effort in dealing with unique issues in statistical classification and feature extraction for remote sensing. Chapter 12 addresses the problem of land-cover map which needs to be updated regularly from the multitemporal image data. The class distribution estimates derived from a supervised training on a previous image of the considered area can represent only rough estimates of the class distributions in the new image to be classified. The training data for the new image is not available in practice. The EM (Expectation-Maximization) algorithm is applied in order to improve such estimates iteratively on the basis of global density function of the new image. The authors also considered classification strategies of using multiple classifier system composed of partially supervised classifiers. The popular nearest neighbor decision rule (NNDR) is closely examined in Chapter 13. The authors carefully point out the shortcomings of the traditional NNDR and presented the concept of Nearest Centroid Neighborhood (NCN) and a k-Near Surrounding Neighborhood (k-NCN) decision rule. The pixel classification of remote sensing problems presents a good data base for comparison of various nearest neighbor decision rules. The chapter also considered combining nearest neighbor classifiers and demonstrated the advantage of using ensemble of classifiers. The Markov random field model of the rich contextual information has been very popular in remote sensing. In Chapter 14, the parameter estimation of Gauss-Markov random field model of hyperspectral images is considered. The resulting large dimensionality is compressed by using the Principal Component Analysis (PCA). The MRF+PCA algorithm clearly demonstrated improved unsupervised classification performance over the use of PCA alone and significant image compression capability.

Sensor fusion has a clear significance in remote sensing, as each sensor has its limited capability making fusion necessary or highly desirable. Chapter 15 is focused on an empirical performance comparison of fusion algorithms for combining the output of segmentation algorithms for PMMW, LADAR reflectance, and LADAR range imagery. Four fusion algorithms along with the three segmentation algorithms are applied to a database containing target instance and target classes. The Bayesian fusion performs the best in terms of jointly optimizing detection and clutter rates.

During the 80's and early 90's there were much excitement on artificial neural networks (or simply called neural networks) which seem to have potential

applications in every area including remote sensing problems. In more recent years there has been better understanding of advantages and limitation of neural networks. Chapter 16 presents the use of competitive neural networks such as the Self-organizing Map for land cover mapping. The authors further employ the Jeffrey-Matusita distance between class pairs for evaluating features. Texture indicators including Energy and Local homogeneity are also considered. The proposed method is shown to perform better than other methods. Another example of making effective use of neural network as presented in Chapter 17 is to structure the input layer in such a way that each pixel belonging to each root window resulting from the contextual feature extraction procedure is represented by an input node. The output of the network, which is usually interpreted to crisp class assignment according to the winner-take-all rule, is softened here considering the value of output neurons directly as classification results, and interpreted as degrees of possibility or compatibility to classes.

Change detection is probably one of the most unique problems in remote sensing, especially considering the need to monitor the state of the Earth surface. The supervised information most likely is not available or high unreliable. Chapter 18 presents two unsupervised change-detection methodologies. The first technique is based on a thresholding algorithm often employed in computer vision. The second technique applies EM algorithm in a transformed space, computed by the Fisher transform. The authors demonstrated the feasibility of the techniques by studying an area damaged by forest fire. There are many other change detection applications such as the study of effect of flood, the change in crops in different months, etc.

A special feature of this book is to include the topic of seismic signal processing which may be considered as remote sensing of the deep earth, rather than the earth surface. All three chapters were presented at the Fourth International Symposium on Computer-Aided Seismic Analysis and Discrimination, held in June 2002 at Westborough, Massachusetts. The meeting was very fortunate to have Dr. Enders A. Robinson, the leader of the field, to give an opening lecture. The full text of his talk is in Chapter 19, "Robust detection using dual sensors". The receiver is made up of dual geophone-hydrophone sensors. The geophone measures the particle velocity while the hydrophone measures the pressure. The receiver is buried at a level below the level of the buried source. This meets the requirement for Einstein deconvolution which can remove all the reverberations and ghosts due to interfaces above the receiver, and unknown source signature in the same operation. The resulting deconvolved seismogram is the desired unit impulse reflection response that would be produced as if there were no layers at all above the buried receiver. The chapter that presents a new approach of Einstein deconvolution is very readable even for non-experts. The second chapter in seismic signal processing,

i.e. Chapter 20, by Dr. Yi Luo and his colleagues at Saudi Aramo presents a new approach of using edge-preserving smoothing (EPS) idea in computer vision to enhance 3-D seismic images. The time slice produced by applying edge detection algorithm with edge-preserving clearly has much more details useful for geophysicists to discover fault zones in oil exploration. The third chapter, i.e. Chapter 21, is another chapter on seismic imaging which develops an inverse scattering algorithm to perform the task of imaging in the absence of accurate velocity information.

The remote sensing problem is not limited to images. Time series analysis is of equal importance. Chapter 22 presents two methods: adaptive Savitzky-Golay filter, and non-linear least-squares fits to asymmetric Gaussian model function, for extracting the seasonality information of the underlying vegetation. The authors applied the methods to NASA/NOAA Pathfinder AVHRR Land Normalized Difference Vegetation Index (NDVI) data over Africa and showed that the two methods complement each other and they may be suitable in different areas depending on the behavior of the NDVI signal.

Image compression techniques for remote sensing data has been an active research area for over forty years, for the objectives of bandwidth reduction and efficient storage. While lossy compression techniques are well documented, the near-lossless compression of the remote-sensing data as presented in Chapter 23 plays a particularly important role in remote sensing. The chapter starts with the review of distortion measures, and then presents the near-lossless image compression through causal and noncausal DPCM for multispectral data, hyperspectral data, and SAR data. Unlike typical lossless compression achieving compression ratios (CR) of two, the methods presented allow a virtually lossless compression with CRs larger than three for 8-bit multispectral data and larger than five for 16-bit hyperspectral data.

The emerging topic of land mine detection or remote sensing just below the earth surface, is presented in Chapter 24, entitled, "Image enhancement in ground penetrating radar (GPR) signal processing to detect landmines". Among many subsurface sensing techniques for detecting and localizing buried landmines, GRS is widely used as it is able to detect both metallic and non-metallic objects and to localize buried objects in a three-dimensional space and obtain discrimination information based on the received signals. However there are clutters in the GPS data, which can come from the reflection by the ground surface and scattering in the soil. The authors present two contrast stretch (histogram manipulation) techniques to enhance the landmine reflected signals. As illustrated by a number of examples, in most cases the task for detecting landmines become simple from enhanced images In other cases post-processing may be needed to further enhance the data.

Infrared technology has been increasingly popular in remote sensing in recent years. Chapter 25 entitled, “Infra-red image processing” presents the use of powerful Pearl’s Bayes networks for extraction of urban regions in a near infra-red image. The approach is equally useful for the texture segmentation of driveable regions for autonomous land vehicles, and for medical applications. The chapter further illustrates that many new techniques in remote sensing information processing are useful in wide range of application domains.

Finally the increasingly important topic of hyperspectral images is examined in Chapter 26, “Hyperspectral imaging (HSI) analysis and applications”. by Dr. Su May Hsu et al. Starting from the hyperspectral imaging sensors, the chapter presents an overview of HSI algorithms for preprocessing, feature extraction and classification & detection. The enormous potential of HSI is illustrated by discussion of HSI application examples such as material identification, anomaly detection, and background classification, along with the issues of HSI fusion with other sensors. The chapter clearly reinforces our view that sensor development in remote sensing must be accompanied by advances in information processing.

Among other topics, knowledge-based method and dedicated software systems are not covered in this volume. While there is continuing progress in developing new information processing techniques for remote sensing, we believe that the book has presented important frontier developments and thus will play a useful role as a resource book for continued progress in remote sensing information processing.

Once again I would like to thank all contributors for their outstanding contributions to the book and their cooperation to meet the manuscript deadline. It was indeed a pleasure to work with them to make this book possible. My special thanks go to Steven Patt, World Scientific Publishing Co. in-house editor of this book, for his efficient effort to make a timely publication of the book possible, and to Mr. Peter Pei-Gee Ho for his help to transmit some chapter manuscripts to the publisher.

C. H. Chen  
March 6, 2003

## CONTENTS

Preface		v
<b>Introduction</b>		<b>1</b>
Chapter 1	Remote Sensing Sensors: Capabilities and Information Processing Requirements <i>John Richards</i>	3
Chapter 2	Transform Methods in Remote Sensing Information Processing <i>C. H. Chen</i>	23
<b>Data Mining</b>		<b>33</b>
Chapter 3	Scene Modeling and Image Mining With A Visual Grammar <i>Selim Aksoy, Carsten Tusk, Krzysztof Koperski and Giovanni Marchisio</i>	35
Chapter 4	A Shape-Based Approach to Change Detection and Information Mining in Remote Sensing <i>Jiang Li, Ram M. Narayanan and Peter Z. Revesz</i>	63
<b>SAR Image Processing</b>		<b>87</b>
Chapter 5	A Review of Polarization Orientation Angle Estimation from Polarimetric Sar Data <i>Jong-Sen Lee, Dale L. Schuler and Thomas L. Ainsworth</i>	89
Chapter 6	Unsupervised Classification of Natural Scenes from Polarimetric Interferometric Sar Data <i>L. Ferro-Famil, E. Pottier and J.S. Lee</i>	105
<b>Wavelet Analysis and Applications</b>		<b>139</b>
Chapter 7	Wavelet Analysis of Satellite Images in Ocean Applications <i>Antony K. Liu, Sunny Y. Wu and Yunhe Zhao</i>	141
Chapter 8	Wavelet-Based SAR Speckle Filters <i>Hua Xie, Leland Pierce and Fawwaz Ulaby</i>	163
Chapter 9	A Wavelet Representation of Multispectral Images <i>Paul Scheunders</i>	197



<b>Military Applications of Remote Sensing</b>	<b>225</b>
Chapter 10 Automating The Estimation of Various Meteorological Parameters Using Satellite Data and Machine Learning Techniques <i>R.L. Bankert, M. Hadjimichael, A.P. Kuciauskas, K.A. Richardson, F.J. Turk and J.D. Hawkins</i>	227
<b>Microwave Remote Sensing</b>	<b>253</b>
Chapter 11 Reconstruction and Resolution Enhancement Techniques for Microwave Sensors <i>David G.Long</i>	255
<b>Statistical Pattern Recognition</b>	<b>283</b>
Chapter 12 Advanced Classification Techniques: Partially Supervised Approaches <i>Lorenzo Bruzzone and Roberto Cossu</i>	285
Chapter 13 Nearest Neighbor Decision Rule for Pixel Classification in Remote Sensing <i>Szymon Grabowski, Adam Jóźwik, C.H. Chen</i>	315
Chapter 14 Unsupervised Segmentation of Hyperspectral Images Using Gauss-Markov Random Fields and PCA <i>C. H. Chen and Hao Chen</i>	329
<b>Automatic Target Segmentation</b>	<b>347</b>
Chapter 15 Multisensor Automatic Target Segmentation <i>Mark R. Stevens and Magnús Snorrason</i>	349
<b>Neural Networks</b>	<b>363</b>
Chapter 16 Category Classification Using Neural Networks <i>Michifumi Yoshioka, Toru Fujinaka and Sigeru Omatsu</i>	365
Chapter 17 Application of Multiresolution Remote Sensing Image Analysis and Neural Techniques to Power Lines Surveillance <i>Elisabetta Binaghi, Ignazio Gallo and Monica Pepe</i>	383
<b>Change Detection</b>	<b>403</b>
Chapter 18 Advances in Unsupervised Change Detection <i>Gabriele Moser, Farid Melgani and Sebastiano B. Serpico</i>	405

<b>Seismic Signal Processing</b>	<b>427</b>
Chapter 19 Remote Detection Using Dual Sensors <i>Enders A. Robinson</i>	429
Chapter 20 Edge-Preserving Smoothing for Enhancing 3-D Seismic Images <i>Yi Luo, Maher Marhoon, Saleh Al-Dossary and Mohammed Alfaraj</i>	455
Chapter 21 Imaging Seismic Reflection Data at The Correct Depth Without Specifying An Accurate Velocity Model: Initial Numerical Examples of An Inverse Scattering Subseries <i>Simon A. Shaw and Authur B. Weglein</i>	469
<b>Time Series Prediction</b>	<b>485</b>
Chapter 22 Seasonality Extraction from Time-Series of Satellite Sensor Data <i>Per Jönsson and Lars Eklundh</i>	487
<b>Image Compression</b>	<b>501</b>
Chapter 23 Near-Lossless Compression of Remote Sensing Data <i>Bruno Aiazzi, Luciano Alparone, Stefano Baronti and Cinzia Lastris</i>	503
<b>Emerging Topics</b>	<b>533</b>
Chapter 24 Image Enhancement in Ground Penetrating Radar Signal Processing to Detect Landmines <i>Xiaoyin Xu and Eric L. Miller</i>	535
Chapter 25 Infra-Red Image Processing <i>Margaret J. Varga and Paul Ducksbury</i>	555
Chapter 26 Hyperspectral Imaging Analysis and Applications <i>Su May Hsu, Hsiao-Hua Burke and Michael K. Griffin</i>	595

# **Introduction**

This page is intentionally left blank

# CHAPTER 1

## REMOTE SENSING SENSORS: CAPABILITIES AND INFORMATION PROCESSING REQUIREMENTS

John Richards

*Research School of Information Sciences and Engineering  
The Australian National University, Canberra ACT 0200, Australia  
Email: john.richards@anu.edu.au*

Following a brief review of developments in optical and radar sensor characteristics, the implications of improved sensor properties for thematic mapping from optical data are discussed. Particular attention is given to the likely poor estimates of second order class statistics obtained with practical training set sizes; methods used to overcome that problem are highlighted. Block diagonalisation of the class covariance matrix is suggested as a viable, practical process to adopt in practice.

Noting the relative maturity of the current generation of sensors, the chapter concludes by commenting on the desirable characteristics of an information system for operational remote sensing purposes, noting especially the need to focus on procedures that manipulate knowledge, rather than data.

### 1. Introduction

From a simple beginning, in which the user community was dependent on essentially a single supply of general purpose satellite image data, we are now in an era:

- (a) where there are a number of competing data suppliers,
- (b) in which spaceborne platforms will possibly be members of a (planned) group of missions, and
- (c) in which missions employ a cluster of sensors, some of which serve a range of applications but with others designed specifically to meet the data requirements of a particular scientific community.

Moreover, data products are now just as likely to be specified regularly in terms of biophysical or geophysical variables<sup>1</sup>, rather than a processing level that describes the integrity of the data product but which nevertheless is still specified in terms of the original measurement variables.

The multiplicity of current programs and sensors means that attention has also shifted from utilising single data products that satisfy the needs of a given application to the use of integrated sets of data that might have come from different



missions and almost certainly recorded with different sensor technologies. Practical and operational information systems now form as much a part of the image interpretation task as processing algorithms themselves.

While many algorithms are mature, the data produced by newer sensors calls into question the efficacy of some of the procedures that have been in use for several decades.

This chapter treats the impact that developments in sensor technology, and notably improvements in spectral resolution and coverage, have had on the data analysis task. We restrict our attention to the challenge of producing reliable thematic maps from recorded data, and thus focus on methods for labelling pixels. More generally this is known as classification. We also restrict ourselves largely to the processing of optical image data, although a brief comment is made on the particularities of radar data when thematic mapping is required.

It is on classification procedures that sensor evolution has perhaps had the greatest influence in respect of image processing. Techniques that performed well on the data generated by early sensors, notwithstanding the limitations inherent in the data themselves, run into trouble when data complexity increases. One would expect that modern sensor data should contain sufficient information for a range of sophisticated applications; however data processing algorithms may often strike mathematical intractabilities or may not be well matched to the complex nature of the data being generated.

Although traditional classification procedures might have to be adapted if they are to be used, the improved characteristics of the data produced by the new generation of sensors permits analytical procedures to be devised that could not be entertained with early data types. Interpretation of hyperspectral data by spectroscopic analysis is one such possibility.

While we concentrate here principally on interpretation and labelling, other image processing procedures, most notably transformation operations, are also affected by increasing sensor sophistication. Comment will be made in passing on how those methods can be modified for use on higher dimensional data sets.

Data compression procedures are also summarised owing to the higher data rates and volumes now being generated.

## **2. The Data Chain Model**

A convenient way to envisage the total problem at hand is to look, first, at a data chain model of a remote sensing system, such as that depicted in Figure 1. Upwelling radiation provides a signal to a transducer or receiver that is converted to data for transmission to the user, often via a ground receiving station and a data processing agent. The data are then analysed, in our case using classification methods, to generate accurate thematic map products that are suited to an application of interest.

Our concentration is on the analysis classification methods themselves and how they have had to be adapted to increasing sensor performance and complexity. It is not just that the improved data types place strains on "tried and true" analytical methods, but the richness of new data sources means that we are challenged to find

methods that allow more meaningful knowledge to be gained about Earth cover types and their usage.

By wishing to understand how sensor development has impacted on analytical procedures and their extensions:

- (a) we need to have some idea of evolution in variety and quality of sensors, and corresponding developments in the nature of the data they produce,
- (b) we need to understand the characteristics and limitations of the labelling methods used successfully with earlier, simpler data types so we can appreciate how they are affected by increasing sensor and data complexity, and
- (c) we need to devise methods for obviating the incipient limitations of standard labelling algorithms, or derive new means for interpretation.

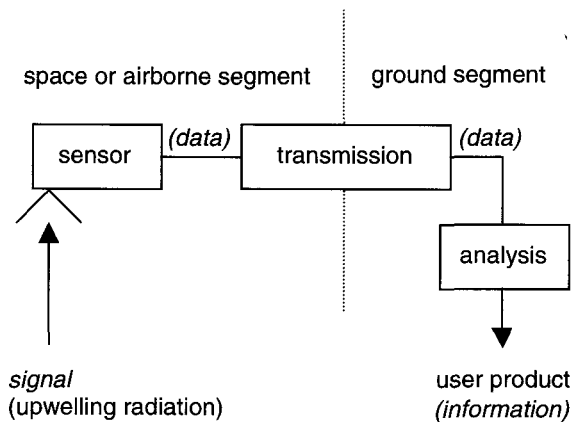


Fig. 1. The data chain in a remote sensing system.

### 3. Remote Sensing Sensor Evolution

The sensor in a remote sensing system is the primary data gathering device from which the recorded signal is made available for subsequent analysis and interpretation. Its design is governed by many factors including, most fundamentally, the primary energy source that forms the basis of the measurement, the applications envisaged for the recorded data and the technologies available. The final specification of the characteristics of a sensor is often determined by a trade-off among those factors. More so now than in the past, sensors may be designed for quite specific applications in mind; in contrast, the sensors of two decades ago were envisaged as serving a wide range of application domains.

The data gathering ability of a sensor is also influenced by the platform on which it is carried. Platform repeat cycle and mission characteristics such as swath width, will impact on recorded data characteristics and will influence data volume

and data rate. Platform velocity, altitude and attitude variations will affect data quality, as will the design specifications of the instruments used.

In a treatment such as this, in which sensors and data processing requirements are considered together, it is important at the outset to establish the range of applications of interest. To this extent, we here exclude sensors used for weather forecasting and climatology. In many ways they are similar in their measurement characteristics to those used for land based remote sensing, being significantly different mainly with respect to spatial resolution. More recent sensors also allow atmospheric properties to be assessed directly but they are beyond the scope of the current treatment. Instead, we focus here on those sensors used to gather information about the surface of the Earth. They therefore have spatial resolutions of better than about 100m x 100m.

### **3.1 Primary energy sources**

There are three primary sources of energy used in remote sensing of the Earth's surface.

The first and most common is the Sun. Sensors optimised to measure sunlight reflected from the surface of the Earth typically operate in a wavelength range between about 0.4 and 2.5  $\mu\text{m}$ , the regime over which the Sun, as an approximation to a black body with a surface temperature of about 6,000K, is significantly emitting energy. By far the majority of sensors used in Earth surface remote sensing operate within these wavelengths, along with thermal sensors and those based on microwave radar technology.

The second principal energy source is the Earth itself. It is also able to be considered in terms of black body radiation theory. At 300K it radiates maximally over the range of about 8 to 12  $\mu\text{m}$ , emitting what we commonly term thermal energy. Sensors designed to operate in those wavelengths in effect measure the temperature (and emissivity) of the Earth itself. Over that range, the reflected solar energy is very small and can be considered insignificant compared with thermal emission.

Along the same lines as thermal emission from the Earth itself is emission from hot bodies such as burning fires. With a temperature of about 1,000K they emit maximally in the range of 3 to 5  $\mu\text{m}$ ; some sensors are designed to operate over that range if the application of interest is, for example, in monitoring forest fires.

The third energy source is that created artificially. In the microwave range, where wavelengths are typically 10,000 times longer than those of light or heat, the energy levels of reflected sunlight and thermal Earth emission are so small that the Earth is for all practical purposes almost dark. That provides the opportunity to employ radar technology to irradiate the Earth artificially with a radio energy source and thereby create an image of Earth surface features in the microwave wavelength range. Known as *active* remote sensing in contrast to *passive*, where

the Sun or Earth provide the principal sources of energy, quite different properties of the Earth's surface can be measured\*.

Using large pixel sizes and sensitive detectors, it is also possible to record self emission of microwave energy from the Earth, allowing thermal measurements to be made under unfavourable weather conditions. Such passive microwave imaging is also useful for soil moisture assessment.

Remote sensing instrumentation is designed to operate with one or more of the three primary energy types. Those developed to measure Earth surface characteristics using reflected solar or emitted thermal energy are frequently called *optical* sensors, in contrast to *microwave* sensors such as imaging radars and microwave radiometers.

### 3.2 *Optical sensor development*<sup>#</sup>

The earliest optical sensors used for remote sensing purposes were carried on aircraft, and it was these that were employed for the first development of automated image analysis tools. Typically they incorporated about twelve visible and infrared channels, with perhaps one or two thermal bands as well.

Among the earliest quasi-operational spaceborne imagery was that obtained with the first Landsat Multispectral Scanner (MSS) in 1972 which recorded data in four very broad wave bands. They were 100 nm in width for the visible green, visible red and first of the near infrared bands and 300 nm for the remaining infrared band. When set against the spectral reflectance characteristic of common Earth surface cover types it is clear that with such broad band definitions very little detail of scientific value can be obtained, although general cover type characteristics can be assessed using appropriate analytical procedures. Nevertheless, in the case of vegetation for example, by measuring the depth of the red absorption band (attributed to chlorophyll) and the magnitude of the infrared response, it is possible to assess the vigour and moisture status of the vegetation in a relative sense and even to do some rudimentary discrimination among vegetation types. But in reality, the spectral resolution of the MSS was simply too coarse for fine differentiation of vegetation and other cover types.

With the availability of the first Landsat Thematic Mapper (TM) data in 1982 the spectral resolution was improved but, more importantly, additional wavebands gave the ability to derive information about cover types that were not revealed well in the MSS bands. Like all spaceborne sensors to that time, however, its design had to be a compromise over application regimes because of the capital cost involved in designing and orbiting such a sensor. In other words, these earlier sensors, like some but not all of their successors, were envisaged as general purpose from the outset.

---

\* Another active remote sensing system is lidar in which laser energy is used to irradiate the Earth. One application of lidar is in bathymetry, in which short wavelength laser energy is used to penetrate shallow oceans and to be reflected off the sea bottom.

<sup>#</sup> For a summary of the technical characteristics of remote sensing instrumentation to about 1996 see Richards and Jia<sup>2</sup>.

The original SPOT HRV sensor, orbited in 1986, while providing only three wave bands, gave a slightly better spatial resolution in its multispectral bands. It also provided the first look at high spatial resolution panchromatic imagery from space. Its application domain was not too different from that of Landsat's TM and MSS with the exception that the Thematic Mapper was better suited to geological applications.

It was not until the advent of hyperspectral sensors, flown first on aircraft platforms and most recently on the Terra<sup>3</sup> and EO spacecraft<sup>4</sup> that sufficient spectral resolution has been made available to allow fine diagnosis of the features of a reflectance spectrum from space. Most hyperspectral sensors operate over the range of wavelengths corresponding to reflected solar energy (0.4 – 2.5  $\mu\text{m}$ ) and include approximately 200 bands in that range with spectral resolutions of about 10 nm. For the first time that suggests that information extraction can be performed based on understanding the spectroscopy of the radiation-matter interaction process, rather than simply by applying classification algorithms.

Table 1. High spatial resolution spaceborne remote sensing missions.  
Panchromatic bands have been ignored in this comparison.

Sensor	First used	Pixel size (m)	Swath (km)	Bands	Bits /pixel	Kbyte/ 100ha
MSS	1972	79	185	4	7, 6 <sup>a</sup>	0.54
TM	1982	30, 120 <sup>b</sup>	185	7	8	6.74
HRV	1986	20	60	3	8	7.5
MESSR	1987	50	100	4	8	1.6
LISS I	1988	73	146	4	7	0.66
LISS II	1991	36	146	4	7	2.7
AVNIR	1996	16	80	4	8	15.63
LISS III	1997	23, 70 <sup>c</sup>	142-146	4	7	5.14
ETM+	1999	30, 60 <sup>b</sup>	185	7	8	6.94
HRVIR	1999	20	60	4	8	10
Ikonos	1999	4	11	4	11	343.75
ASTER	1999	15,30 <sup>c</sup> ,90 <sup>b</sup>	60	15	8,8 <sup>c</sup> ,12 <sup>b</sup>	25.37
Hyperion	2000	30	7.7	22	12	366.67
ALI	2000	30	37	10	8	11.11

a For the second near infrared band

b For the thermal band(s)

c For the middle infrared band(s)

Table 1 shows the characteristics of a number of spaceborne optical sensors. Only those with high spatial resolution (pixel size better than about 100m x 100m) are included. Also, the band information provided does not cover any panchromatic bands; that data is more of value to topographic and cartographic purposes than to thematic mapping, which is the subject of this treatment. It could be noted in passing, though, that high resolution panchromatic data is often used to "sharpen"



colour products created from lower resolution multispectral data when visual interpretation is envisaged.

Also shown in Table 1 is the maximum data volume produced for 100ha of ground coverage. This is used to illustrate, in principle, the increasing volume of data to be processed with improvements in sensor technology. When plotted logarithmically, as shown in Figure 2, it is suggested that the maximum demand for data processing (measured by slope in the maximum data volume “straight line” envelope shown) is increasing by about an order of magnitude per decade. Clearly this is a very coarse measure, since it assumes all bands for each sensor would be employed in a given classification exercise and that those sensors with the maximum data generation would be chosen in preference to the others. Task-specific sensor selection, and the use of feature reduction techniques, makes both unlikely. Nevertheless the measure highlights the likely increase in data to be analysed. It could be noted that the *average* increase in data processing demand is about an half an order of magnitude (i.e. about 3 times) per decade.

Clearly, improvements in computational technology have more than kept pace with the increase in data volume, but algorithm development may have not, as we will illustrate below.

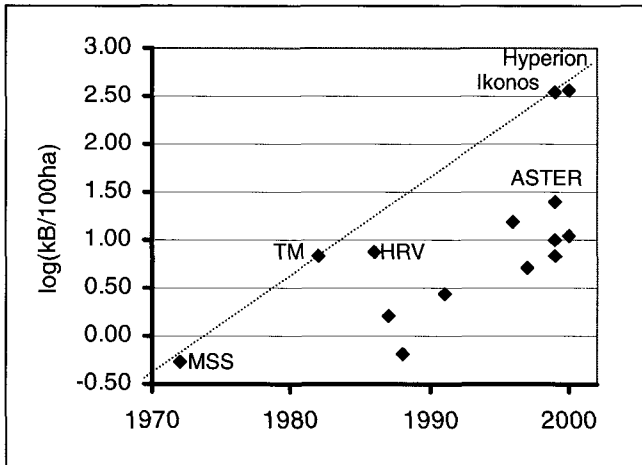


Fig. 2. Maximum data volume generated per 100ha for the high spatial resolution multispectral sensors in Table 1.

Interestingly, the maximum envelope in Figure 2 is defined by the EOS sensor family (MSS, TM and Hyperion). Given that the platform repeat cycles are the same (approximately so for MSS), then their data rates are in the same proportions. These are data rates in terms of analysis demand not the telemetry rates used for transmission from orbit to a receiving station.

### 3.3 Radar sensor development<sup>2</sup>

As with optical data capture, the first imaging radar systems were airborne and, in fact, pre-dated optical airborne systems. The first spaceborne radar was that carried on the Seasat satellite launched in 1978, followed by the shuttle-based SIR-A mission in 1981. The radars on both Seasat and SIR-A recorded just a single channel of microwave image data. With the launch of SIR-B in 1984, a variable look angle was introduced making possible, for the first time from a spaceborne platform, the creation of a multiply dimensioned image, characterised by using different look angles for the same region being imaged.

Dimensionality was increased further with the shuttle SIR-C/X-SAR mission flown in 1994 through the use of multiple polarisations, multiple wavelengths and multiple incidence angles.

The European and Japanese free flying radar satellites ERS-1 and JERS-1, first launched in 1991 and 1992 respectively, ushered in the era of operational radar imaging from space. Their instruments again were single dimensional. Radarsat-1 launched in 1995 provided a variable look angle and thus increased data dimensionality.

While with optical systems, data complexity has increased through improvements in spectral resolution (and number of channels), complexity in spaceborne radar systems has resulted from increased wavelengths (analogous to bands for optical systems), increased polarisations (and, because of the coherent nature of the radiation, the ability to synthesise other polarisations), and the use of variable look angles.

As a result of the use of coherent radiation in imaging radar systems, interferometric imaging is possible, permitting fine topographic detail on the Earth's surface to be measured. Although demonstrated initially with data from general purpose imaging radar missions<sup>5</sup>, this development has led to the dedicated Shuttle Radar Topography Mission for wide scale topographic mapping of the Earth's surface<sup>6</sup>.

## 4. The Data Processing Problem

The intersection between the sensors themselves and the demand they place on data processing for information extraction lies, first, in the volume of data they produce, as suggested in Figure 2. This is determined by their characteristics and specifically the number of separate bands of data recorded, the pixel rate (sometimes expressed as number of pixels per frame) and the number of brightness or radiometric values associated with the measurements represented by each pixel in each band.

For a given application not all of the data produced by a particular sensor would need to be used. Instead, sensors are often designed to meet the needs of a variety of applications and thus cover spectral ranges and radiometric values beyond those required for a specific application. Nevertheless, as noted above, the total data volume generated by a sensor is a useful primary indicator of the demands made on image interpretation algorithms.

Processing *demand* as expressed by the volume of data to be handled is only one consideration. *Complexity* in data processing also arises with new sensors because of the improved nature of the recorded data and the type of information that the user is, in principle, now able to extract.

Pioneering work in the development and application of computer assisted procedures for creating thematic maps from remotely sensed data was carried out at the Laboratory for Applications of Remote Sensing (LARS) at Purdue University in the 1960s and 1970s<sup>7</sup>. It was realised, by association with signal processing, that maximum likelihood estimation techniques should perform well with the nature of the multispectral data recorded by aircraft sensors and later by the then new Landsat Multispectral Scanner data.

The essential objective of the analysis has remained the same since that time - it's just that the complexity of the task has increased. In summary, to produce a thematic map from an image we need to label each pixel using a decision rule along the lines:

$$x \in \omega_i \text{ if } g_i(x) > g_j(x) \text{ for all } j \neq i \quad (1)$$

where  $x$  is a vector of the multispectral measurements (brightness values) for the pixel being labelled and  $\omega_i \ i=1, \dots, C$  is the  $i^{\text{th}}$  class label of  $C$  total possible labels.

$g_i(x)$  is a discriminant function, which can be defined in many ways, depending on the particular algorithm or technique adopted for the classification task<sup>2</sup>. When maximum likelihood estimation is used, it is assumed that the pixels in each class are distributed in the spectral domain in a multidimensional normal fashion - i.e. a Gaussian distribution model is assumed - so that the discriminant functions take the form<sup>2</sup>:

$$g_i(x) = \ln p(\omega_i) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (x - m_i)' \Sigma_i^{-1} (x - m_i) \quad (2)$$

in which  $m_i$  is the mean brightness vector for class  $\omega_i$  and  $\Sigma_i$  is its covariance matrix;  $p(\omega_i)$  is the prior probability of the class, which is the likelihood with which  $\omega_i$  could be guessed as the correct class label for the pixel in the absence of the remotely sensed information.

Although ostensibly simple to apply, the maximum likelihood technique has pitfalls if the user is not careful in its application, and particularly in understanding the limitations of the algorithm. The first relates to the assumption of class normality. It was recognised early on that this is essentially an incorrect assumption for most recorded data. Rarely do pixels from individual classes distribute themselves spectrally in a Gaussian fashion, even if they do form clusters in the spectral domain. More often they are part of a continuum<sup>8</sup>. Nevertheless, careful use of clustering methods beforehand<sup>9</sup> allows the data to be conditioned in such a manner that maximum likelihood classification works well, leading to its being the most commonly adopted labelling technique applied to spectral data.

A second limitation with maximum likelihood estimation was also understood and overcome at the outset, at least with the limited number of spectral channels then available; that has to do with over-dimensioned data. In order to use maximum

likelihood classification the mean vector and covariance matrix  $\mathbf{m}_i$  and  $\Sigma_i$  for each class have to be known. They are found by the analyst having identified previously a set of pixels known to come from each of the classes of interest. Using those prototype - or *training* - pixels, values for  $\mathbf{m}_i$  and  $\Sigma_i$  are estimated. For  $N$  dimensional data (i.e.  $N$  multispectral bands) there are  $N$  components of  $\mathbf{m}_i$ .  $\Sigma_i$  is an  $N \times N$  matrix, but there are only  $\frac{1}{2}N(N+1)$  distinct components, since the matrix is symmetric.

In principle, the user should need only to identify around  $\frac{1}{2}N(N+1)$  pixels for each class in order to estimate the class parameters. For Landsat MSS data that would suggest 10. However, the way the training pixels are chosen in practice does not ensure that they are independent samples; more than this number is needed to given reliable estimate of the class statistics<sup>2,7</sup>. Should not enough independent samples be chosen or available then the statistics will be poorly estimated and the labelling of a given image will have low accuracy.

In the remote sensing literature this effect is often referred to as the Hughes phenomenon<sup>7</sup>, and it is posed in the following manner. For a given set of training pixels the accuracy of a classification will generally increase as the number of bands available (i.e. dimensionality) increases; there will however be a dimensionality beyond which the accuracy of the classification outcome will deteriorate. To avoid reaching that stage it has been suggested that between 10N and 100N training pixels per class be used<sup>7</sup>.

It is this problem that bedevils the application of maximum likelihood methods to modern sensors.

## 5. Classification with Increased Sensor Complexity

### 5.1 *The problem of limited training data*

From a classifier performance point of view, the most significant development in sensor characteristics has been the much greater number of bands available, especially with hyperspectral instruments such as MODIS and Hyperion. Not only does that lead to an increased classification time for maximum likelihood classification, owing to the quadratic dependence on dimensionality of the discriminant function in (2), but often it is not possible to find enough training samples per class to avoid poor estimation of the elements of the covariance matrix. In other words, there is a major consideration with respect to the reliable determination of class statistics for hyperspectral data sets. As a guide, between about 2000 and 12,000 pixels per class should be used with Hyperion data if all channels were to be used in labelling; clearly that is impracticable.

### 5.2 *Feature reduction*

A rather logical approach to solving the problem has been to seek to reduce data dimensionality prior to the application of classification algorithms, particularly since it is expected that for many application domains the highly dimensioned hyperspectral data sets would contain major redundancies. For example, principal

components transformations have been used<sup>10</sup>. However, as correctly pointed out by Morgan et al<sup>11</sup>, the principal components transformation is sub-optimal from a labelling standpoint because the correlation that is exploited for dimensionality reduction is that of the global data set, and is not class-specific. Nevertheless, the transformation frequently does perform well for feature reduction, particularly for highly correlated data sets, because the classes themselves are often distributed approximately along the diagonals in multispectral space. A fast principal components transformation for hyperspectral data, which benefits from segmenting the covariance matrix in the manner described in Section 5.4 below, has been proposed by Jia and Richards<sup>12</sup>.

The problem with trying to apply more appropriate feature selection methods to hyperspectral data, such as those based on any of canonical analysis, divergence and Jeffries Matusita distance<sup>2,7</sup>, is that they require class covariance matrices to be available. Thus we have the same difficulty as in attempting classification in the first place.

Lee and Landgrebe<sup>13</sup> have devised a feature reduction technique in which decision boundaries are located between training set pixels directly and redundant features - i.e. those that do not contribute significantly to discrimination - are transformed out. The remaining features then form the reduced set for classification purposes. The procedure is known as Decision Boundary Feature Extraction (DBFE).

An alternative is Discriminant Analysis Feature Extraction<sup>14</sup> (DAFE). In essence it is similar to canonical analysis<sup>2</sup> and, as with the traditional separability measures, requires class covariance information to be estimated beforehand. This problem is overcome in the Non-parametric Weighted Feature Extraction procedure proposed recently<sup>15</sup>.

### ***5.3 Non-parametric labelling methods***

Non-parametric, non-statistical schemes have been explored in order to obviate the difficulties with covariance estimation altogether. These have included the use of Support Vector Machines (SVM), which are a practical extension of the linear learning machine (Perceptron) concept<sup>16</sup> known for many decades, but in which the training process is guided by only the most significant training pixels (support vectors) and in which linear inseparability is handled through expression of the pixel vectors in a higher dimensional space.

In the Perceptron, two class training is achieved by setting up a linear hyperplane to separate all pixels into one of the two classes by using each pixel's position in the spectral domain relative to the others. Essentially the hyperplane is found by iterating an initial guess into a final state by reference to *all* the training pixels. In the SVM only those pixels "closest" to the separating hyperplane are needed and used to determine its position. Further, the position of the hyperplane is so determined that the distance (margin) between the closest pixels in each of the two classes and the hyperplane position is maximised. If the training data sets are linearly inseparable, which is highly likely for real world applications, a transformation to a higher dimensional space is used to force separability.

Being non-parametric, the SVM does not suffer the parameter estimation difficulty of maximum likelihood methods, although its computational demand can be high, and will increase linearly with data dimensionality.

Recent experiments with SVM have shown it to perform well<sup>17,18</sup> and, for the limited trials reported, it performs better than neural networks and Gaussian maximum likelihood. The poorer performance of the maximum likelihood algorithm is not surprising, given the parameter estimation difficulties already mentioned.

#### 5.4 Modifying the maximum likelihood rule

Another solution has been to stay with the traditional maximum likelihood classification approach for thematic mapping, but devise procedures that permit reasonable estimation of the covariance matrix, particularly for classes for which limited training samples are available. An immediate possible benefit of this approach is that standard classification software, based on class normality and the long-standing maximum likelihood rule, can be retained.

Generally there have been two approaches to this particular problem. In the first, approximation methods are employed to give covariance elements that are sufficiently accurate that classifier performance does not unreasonably suffer. In the second, the properties of the actual covariance matrix revealed with real data sets are exploited to reduce the problem to one of smaller dimensionality.

Kuo and Landgrebe<sup>19</sup> review and improve on several schemes that seek to determine reliable versions of the class conditional sample covariance matrix by using mixtures of the actual sample covariance matrix, estimated from the available training data, and the pooled covariance matrix computed over all the training samples. Properties of these matrices, including their traces and diagonal forms, are sometimes used in these regularisation procedures.

The maximum likelihood rule has been iterated in Jackson and Landgrebe<sup>20</sup> to improve covariance estimates and thus labelling accuracy. By using limited available training data, an initial classification is performed. This is known to be affected by poorly estimated class statistics. Previously unknown pixels that have now been labelled in this classification step are then used with the original training data (although weighted down in their influence) to give better covariance estimates. The classification is then repeated. Iterating through this sequence produces good results in relatively few iterations provided the initial statistics estimates and labelling accuracy are not too poor. The procedure is improved if acceptable covariance estimators, designed specifically for use small training sets, are also employed<sup>21</sup>.

As an alternative, consider the actual properties of the covariance matrix from a real data set. In the following, we are going to be guided by the *correlations* among the bands of data and so examine the correlation matrix rather than the covariance matrix. Even so, the properties to be adduced will provide guidance on how the covariance matrix can be handled.

Figure 3(a) shows the correlation matrix (in pseudo-image form) for a set of pixels recorded by the AVIRIS sensor; 196 wavebands have been used, with white implying 100% positive or negative correlation and black indicating zero

correlation<sup>22</sup>. What is immediately obvious is the dark stripes running through the correlation matrix corresponding to the water absorption bands in the recorded spectral data. Those bands can be used to segment the correlation (and thus the covariance) matrix into blocks. Figure 3b, shows such a segmentation. The blocks along the diagonal generally show the highest correlations, while those in the off-diagonal positions often show lower correlations and in Figure 3(b) are set to zero. As might be expected this behaviour suggests that adjacent bands are the most strongly correlated, while those further apart in the spectrum generally show lower correlations. A notable exception, however, can be the correlations among the middle infrared bands either side of the 1.9  $\mu\text{m}$  water absorption.

We can, therefore, with little loss of covariance information, approximate the correlation and covariance matrices just by the block diagonal form depicted in Figure 3(b) and can use the properties of block diagonal matrices\* to simplify the classification task. For example, it can be shown that<sup>22</sup>

$$|\Sigma_i| = \prod_{k=1}^K |\Sigma_{ik}|$$

where  $|\Sigma_{ik}|$  is the determinant of the covariance matrix of the  $k^{\text{th}}$  diagonal block, of the total  $K$  blocks in the full class covariance matrix  $\Sigma_i$ . Similarly, it can be demonstrated that

$$\Sigma_i^{-1} = \sum_{k=1}^K \Sigma_{ik}^{-1}$$

where  $\Sigma_{ik}^{-1}$  is the inverse of the covariance matrix of the  $k^{\text{th}}$  block for the  $i^{\text{th}}$  class.

With these the discriminant function in (2) can be expressed

$$g_i(\mathbf{x}) = \sum_{k=1}^K g_{ik}(\mathbf{x})$$

The discriminant function of (2) therefore becomes the sum of the discriminant functions computed over the individual blocks. The largest block of the covariance matrix determines the number of training pixels needed to avoid the Hughes phenomenon and thus poor estimates of the components of the covariance. Typically this block will be no larger than about one third of the size of the complete covariance matrix. The discriminant function will also be faster to evaluate owing to the quadratic time saving.

Using the positions of the water absorption bands to define the block boundaries is a fairly crude tool, and other segmentations of the covariance matrix might be better, as shown by Jia<sup>22</sup>. For example, Figure 3(c) shows a segmentation in which

---

\* Also called quasi-diagonal matrices.

finer block diagonalisation is realised by making use of obvious structure in addition to that provided by the absorption bands. Although sub-optimal, using the locations of the water absorption bands is nevertheless a simple, repeatable and data-independent method for partitioning the matrix and is recommended. Work has yet to be undertaken to test the penalty incurred in classification through making this simplification.

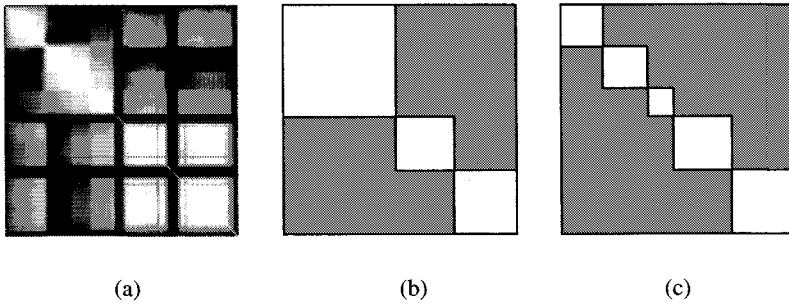


Fig. 3. (a) Covariance matrix of a 192 band image recorded by AVIRIS. (b) Band diagonal form based on water absorption bands. In the white blocks the original covariance information would be retained, whereas in the grey regions the correlations (and thus covariances) are assumed to be zero. (c) Band diagonal form with further sub-division within the visible and near infrared region.

A benefit of this approach is that standard image classification software can be used. All that is needed is a means for aggregating the individual discriminant functions for each class.

Full details of the block diagonalisation approach are given in Jia<sup>22</sup>, and Jia and Richards<sup>23</sup>.

Roger<sup>24</sup> also examines the properties of the covariance matrix, although in its inverse form. He shows that its sparseness allows simplified maximum likelihood procedures to be devised by approximating the inverse matrix by a band diagonal form, as against the block diagonal form discussed here.

### 5.5 Other information extraction techniques for hyperspectral data sets

As a result of the fine spectral resolution and large number of bands available in hyperspectral data it is possible to apply spectroscopic knowledge to features observed in a recorded spectrum. It is also possible to compare the reflectance spectrum for a pixel recorded by a remote sensing program with those of known cover types recorded by a laboratory-based spectrometer, in order to find a match for the remotely sensed spectrum.

We can also seek to un-mix the pixels in the data. The problem of mixed pixels - i.e. those that do not consist of a single cover type - has been a long-standing one in remote sensing data capture. The quality of data now available means that it is possible to create reasonable models that estimate the proportions of "pure" cover



types within a given recorded pixel. A good overview of unmixing is given in Keshava and Mustard<sup>25</sup>.

## 6. Impact of Sensor Evolution on other Image Processing Operations

Even though we are concentrating here on labelling processes, it is nevertheless important to comment on the effect new sensor technologies have on other image processing operations. In many cases the problem is simply one of scaling the computational demand through the added channels of data. But for operations involving the use of second order statistics, the same problem arises as when employing maximum likelihood classification. Thus, separability assessment using divergence-based techniques, canonical analysis for dimensionality reduction and, to a lesser extent, principal components transformation all suffer. Block diagonalisation of the covariance is as useful a remedy here as it is for classification. The reason that principal components analysis is not so badly affected is that it does not rely upon the estimation of class specific covariance matrices, but rather on the global covariance of a large data set, for which sufficient pixels are usually available.

## 7. Labelling from Radar Data

Although beyond the scope of this treatment it is important to comment on the interpretation of radar imagery since, even with single band radar data, pattern classification procedures don't work particularly well. There are several reasons for this. First, the speckle noise present in radar (for which the signal noise ratio is often not much better than 0 dB) means that variance resulting from cover type differences is usually dominated by speckle variance. In order to derive any meaningful labelling it is often necessary to average over large numbers of pixels to reduce the speckle noise more than might already have been the case through what is called look-averaging. This suggests that spatial resolution needs to be sacrificed for labelling accuracy.

Secondly, the nature of the mechanisms by which cover types appear in radar imagery means that application of discriminant function based techniques will often be unsuccessful. For example, while most reflected energy in optical remote sensing is the result of surface scattering, for radar imagery there are also major components of volume scattering in a typical scene; there are also strong scattering elements such as trees and buildings that act as corner reflectors. Pixels of comparable brightness might have resulted from any of these mechanisms. There is also the prospect, particularly with long wavelength radars, for sub-surface reflection in very dry regions; this can complicate the interpretation process further.

A viable approach for the interpretation of radar data is to model the energy-matter interactions that lead to the recorded radar data and then invert those models to derive properties of interest.

Lee and Grunes<sup>26</sup> summarise recent approaches used for labelling radar data, including for multi-band and multi-polarization imagery. Interestingly, the Support

Vector Machine approach has also been used with multi-band, multi-polarisation aircraft radar data with apparently good results, although the image was spatially pre-filtered to reduce the effect of speckle<sup>27</sup>.

## 8. Data Compression

Compression of data for transmission and storage purposes is a standard signal processing procedure and has been used widely for data telemetry since the earliest days of remote sensing. With the increasing data rates from the more recent sensors evident in Figure 2, the need for efficient coding schemes for compression has become more critical. Schemes have ranged from the use of principal components<sup>28</sup> (as a lossy procedure) through to adaptation of vector coding techniques used in the television industry<sup>29,30</sup>. Qian et al<sup>31</sup> have recently proposed a vector coding technique in which the code books are significantly smaller than with the conventional method and which seems to offer prospects for compressions to levels of 30:1 while maintaining sufficient data integrity for most applications.

## 9. Data Fusion – Towards Operational Classification and the Knowledge Chain Model in Remote Sensing

Many of the contemporary sensors in remote sensing are designed to operate as part of an information system. Each provides information on a particular aspect of the Earth or one of its primary cycles, and *together* they provide a synergy of data that allows knowledge to be built up about fundamental processes, land cover and land use in manner not possible with sensors operating independently.

Deriving viable techniques for merging data has therefore been a subject of considerable attention in the field of image classification for many years. Twenty years ago, effort was focussed on simple problems, such as incorporating the effect of spatial context or topographic information, into a classification<sup>32, 33</sup> whereas recently the more general problem of labelling pixels described by several separate sets of data has been of interest. Techniques have ranged from the use of evidential theory<sup>34</sup> to the adoption of expert system methods<sup>35</sup>.

Generally, the employment of several different sources of data together to form joint inferences is called *data fusion*. In an operational context however, one must question whether fusion at the data level is practical. Given that data sources may not all be available at the same time, that some may already have been transformed from data to biophysical variables and that there is value in performing labelling in the context of a geographical information system, where individual data types may be held and maintained by the primary data agencies rather than by the users themselves, a more pragmatic approach is to fuse at the level of labels<sup>36</sup> or decisions<sup>37</sup>. Figure 4 illustrates this concept, depicting the data chain model of Figure 1 as (at least) one component of a data/knowledge system.

In label level fusion decisions can be taken about the correct class for a pixel from each available data source individually, using analytical techniques optimally matched to the characteristics of those sources. A joint inference about the pixel is

then formed by operating on the information labels generated from the individual analyses. This approach seems to offer most prospect of operationalising a distributed data, information and knowledge processing system to meet the needs of contemporary sensors ranging from the general purpose to those designed with specific applications in mind.

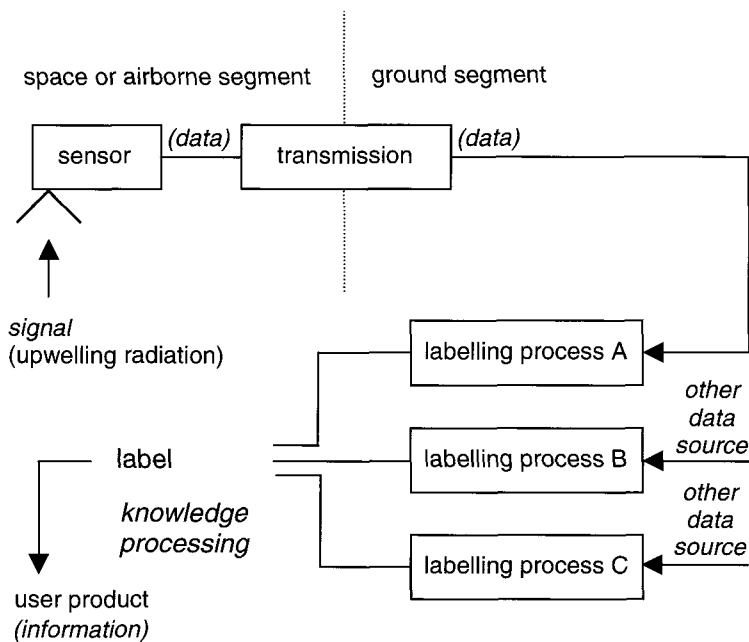


Fig. 4. The data chain as a knowledge chain.

This is not to say that the challenge of devising viable image interpretation tools for the data sets is not important; rather, it means that attention can be focussed on deriving the most suitable analytical procedures for each data source, independently of the other types of data that will be used with it to achieve a final outcome.

## 10. Concluding Comments

The move towards special purpose instruments, instrument clusters and multi-platform missions will undoubtedly continue, as will the trend to specify recorded data in terms of those variables most familiar to the user. Indeed, with an appropriately developed information system and suitable analytical tools, many future end-users might not even be aware that the primary source of data gathered for their particular application was from a spaceborne platform. If that can be achieved we will know that remote sensing has, in a sense, become properly operational and we would expect to see a user uptake of the technology long hoped for but not really experienced in the past three decades. While mission design,

because of capital costs, will always involve some form of compromise, current and anticipated data types and their frequencies are well suited to a very wide range of user needs so that such an operational scenario should not be too far off, once the interpretation tools needed for the newer sensor data sets are finally settled.

## References

1. Justice, C.O., Townshend, J.R.G., Vermote, E.F., Masuoka, E., Wolfe, R.E., Saleous, N., Roy, D.P. and Morisette J.T. (2002). An overview of MODIS Land data processing and product status, *Remote Sensing of Environment*, 83, pp. 3-15.
2. Richards, J.A. and Jia, X. (1999). *Remote Sensing Digital Image Analysis, 3<sup>rd</sup> Ed.*, Springer, Berlin.
3. <http://terra.nasa.gov>
4. Ungar, S.G. (2001). Overview of EO-1: the First 120 days. *Proc. Int. Geoscience and Remote Sensing Symposium (IGARSS) 2001*, The University of New South Wales, Sydney, 9-13 July.
5. Zebker, H.A. and Goldstein, R.M. (1986). Topographic mapping from interferometric synthetic aperture radar observations. *J. Geophysical Res.*, 91, No B5, pp. 4993-4999.
6. <http://www.jpl.nasa.gov/srtn/>
7. Swain, P.H. and Davis, S.M. (1978). *Remote Sensing: the Quantitative Approach*, McGraw-Hill, N.Y.
8. Richards, J.A. and Kelly, D.J. (1984). On the concept of spectral class. *Remote Sensing Letters*, 5, pp. 987-991.
9. Fleming, M.D., Berkebile, J.S. and Hofer, R.M. (1975). Computer aided analysis of Landsat-1 MSS data: A comparison of three approaches including a modified clustering approach. *Information Note 072475*, Laboratory for Applications of Remote Sensing, West Lafayette, Indiana.
10. Liew, S.C., Chang, C.W. and Lim, K.H. (2002). Hyperspectral land cover classification of EO-1 Hyperion data by principal component analysis and pixel unmixing. *Proc. Int. Geoscience and Remote Sensing Symposium (IGARSS) 2002*, Toronto, 24-28 June.
11. Morgan, J.T., Henneguella, A., Crawford, M.M., Ghosh, J and Neuenschwander, A. (2002). Best bases Bayesian hierarchical classifier for hyperspectral data analysis. *Proc. Int. Geoscience and Remote Sensing Symposium (IGARSS) 2002*, Toronto, 24-28 June.
12. Jia, X. and Richards, J.A. (1999). Segmented principal components transformation for efficient hyperspectral remote sensing image display and classification. *IEEE Transactions on Geoscience and Remote Sensing*, 37, pp. 538-542.
13. Lee, C and Landgrebe, D.A. (1993) Feature extraction based on decision boundaries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15, pp. 388-400.
14. Landgrebe, D.A. (1999). Information extraction principles and methods for multispectral and hyperspectral image data. In Chen, C.H. (Ed) *Information Processing for Remote Sensing*, World Scientific, Singapore. pp. 3-37.
15. Kuo, B-C and Landgrebe, D.A. (2002). Hyperspectral data classification using non-parametric weighted feature extraction. *Proc. Int. Geoscience and Remote Sensing Symposium (IGARSS) 2002*, Toronto, 24-28 June.
16. Nilsson, N.J. (1965). *Learning Machines*, McGraw-Hill, N.Y.
17. Lennon, M, Mercier, G. and Hubert-Moy, L. (2002). Classification of hyperspectral images with nonlinear filtering and support vector machines. *Proc. Int. Geoscience and Remote Sensing Symposium (IGARSS) 2002*, Toronto, 24-28 June.

18. Melgani, F. and Bruzzone, L. (2002). Support vector machines for classification of hyperspectral remote-sensing images. *Proc. Int. Geoscience and Remote Sensing Symposium (IGARSS) 2002*, Toronto, 24-28 June
19. Kou, B-C. and Landgrebe, D.A. (2002). A covariance estimator for small sample size problems and its application to feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 40, pp. 814-819.
20. Jackson, Q and Landgrebe, D.A. (2001). An adaptive classifier design for high-dimensional data analysis with a limited training data set. *IEEE Transactions on Geoscience and Remote Sensing*, 39, pp. 2664-2679.
21. Jackson, Q and Landgrebe, D.A. (2002). An adaptive method for combined covariance estimation and classification. *IEEE Transactions on Geoscience and Remote Sensing*, 40, pp. 1082-1087.
22. Jia, X. (1996). *Classification of Hyperspectral Data Sets*. PhD Thesis, The University of New South Wales, Sydney.
23. Jia, X. & Richards, J.A. (1994). Efficient maximum likelihood classification for imaging spectrometer data sets. *IEEE Transactions on Geoscience and Remote Sensing*, 32, pp. 274-281.
24. Roger, R.E. (1996). Sparse inverse covariance matrices and efficient maximum likelihood classification of hyperspectral data. *Int. J. Remote Sensing*, 17, 589-613.
25. Kirshava, N. and Mustard, J. F. (2002). Spectral unmixing, *IEEE Signal Processing Magazine*, Vol 19, No 1, January.
26. Lee, J-S and Grunes, M.R. (1999). Polarimetric SAR speckle filtering and terrain classification – an overview. In Chen, C.H. (Ed) *Information Processing for Remote Sensing*, World Scientific, Singapore. pp. 113-138.
27. Fukuda, S and Hirosawa, H. (2001). Support vector machine classification of land cover: application to polarimetric SAR data. *Proc. Int. Geoscience and Remote Sensing Symposium (IGARSS) 2001*, The University of New South Wales, Sydney, 9-13 July.
28. Lim, S., Sohn, K.H. and Lee, C. (2001). Principal component analysis for compression of hyperspectral images. *Proc. Int. Geoscience and Remote Sensing Symposium (IGARSS) 2001*, The University of New South Wales, Sydney, 9-13 July.
29. Ryan, M.J., and Arnold, J.F. (1997). Lossy compression of hyperspectral data using vector quantisation. *Remote Sensing of Environment*, 61, pp. 419-436.
30. Ryan, M.J and Arnold, J.F. (1997). The lossless compression of AVIRIS images by vector quantisation. *IEEE Transactions on Geoscience and Remote Sensing*, 35, pp. 546-550.
31. Qian, S-E., Hu, B., Bergeron, M., Hollinger, A. and Oswald, P. (2002). Quantitative evaluation of hyperspectral data compressed by near lossless onboard compression techniques. *Proc. Int. Geoscience and Remote Sensing Symposium (IGARSS) 2002*, Toronto, 24-28 June.
32. Kettig, R.L. and Landgrebe, D.A. (1982). Classification of multispectral image data by extraction and classification of homogeneous objects. *IEEE Transactions on Geoscience Electronics*, GE-14, 1976, 19-26.
33. Richards, J.A., Landgrebe, D.A. and Swain, P.H. (1982). A means of utilizing ancillary information in multispectral classification. *Remote Sensing of Environment*, 12, pp. 463-477.
34. Moon, W.L. (1990). Integration of geophysical and geological data using evidential belief functions. *IEEE Transactions on Geoscience and Remote Sensing*, 28, pp. 711-720.
35. Srinivasan, A. and Richards, J.A. (1993). Analysis of GIS spatial data using knowledge-based methods. *Int. J. of Geographic Information Systems*, 7, pp. 479-500.

36. Richards, J.A. (2001). Analysis of mixed data types: the unsolved problem. *Conference on Image and Signal Processing for Remote Sensing VII, 8<sup>th</sup> International Symposium of Remote Sensing*, 17-21, Toulouse, September.
37. Solaiman, B. (1999). Information fusion for multispectral image classification post-processing. In Chen, C.H. (Ed) *Information Processing for Remote Sensing*, World Scientific, Singapore. pp. 339-367.

# CHAPTER 2

## TRANSFORM METHODS IN REMOTE SENSING INFORMATION PROCESSING

C. H. Chen

*University of Massachusetts Dartmouth, N. Dartmouth, MA 02747 USA*  
*E-mail: cchen@umassd.edu*

One unique feature in the remote sensing problems is that a significant amount of data are available, from which desired information must be extracted. Transform methods offer effective procedures to derive the most significant information for further processing or human interpretation and to extract important features for pattern classification.

In this chapter a survey of the use of major transforms in remote sensing is presented. These transforms have significant effects on data reduction and compression and on pattern recognition as features derived from orthogonal or related transforms tend to be very effective for classification. After the introduction, we will examine the PCA and discriminant analysis transform, empirical orthogonal functions (EOF), component analysis and an independent component analysis (ICA) algorithm, followed by concluding remarks.

### 1. Introduction

In remote sensing, the large amount of data makes it necessary to perform some type of transforms that preserve the essential information while considerably reducing the amount of data. In fact most remote sensing image data are redundant, dependent and noisy. Transform methods can help in three ways: effective data representation, effective feature extraction, and effective image compression.

Orthogonal transforms are highly desired as they tend to extract “non-overlapped” information from the data and thus result in more efficient representation. These transforms generally involve linear operations. Let  $\mathbf{x}$  be an input vector and  $\mathbf{y} = \mathbf{W}\mathbf{x}$  be an output vector with reduced dimensionality after applying a transformation represented by the matrix  $\mathbf{W}$ . For orthogonal transform, the vectors forming  $\mathbf{W}$  are orthogonal. The optimal orthogonal transform in the minimum mean square error sense is the discrete Karhunen-Loeve transform (KLT). It has several equivalent names such as the empirical orthogonal function (EOF) and principal component analysis (PCA). The topic of PCA for remote sensing has been well examined in the literature (see e.g. [1,2,3]). For the purpose of pattern classification, such transforms can be useful in feature extraction. However the low mean square error does not guarantee the minimum probability of error typically used as a classification performance measure. In fact the transforms for data representation are generally not the same as those for discrimination.

In this chapter we will examine the PCA and discriminant analysis, empirical orthogonal functions (EOP), component analysis, independent component analysis (ICA) and an ICA algorithm. The topics of wavelet transform and image compression are treated in detail in other chapters of this volume. The references listed though incomplete are representative of the activity in the topic areas.

## 2. PCA and Discriminant Analysis

Consider a set of  $N$   $n$ -dimensional vectors  $\{X_k, k = 1, 2, \dots, N\}$  which are mapped onto an  $m$ -dimensional feature space, where  $m < n$ . The new feature vectors  $Y_k$  are defined as,

$$Y_k = W^T (X_k - E(X_k)) \quad , \quad k = 1, 2, \dots, N$$

where  $W$  is a matrix of eigenvectors. Let  $S_T, S_b$  and  $S_w$  respectively the total scatter matrix, the between-class and within-class scatter matrices of  $X_k$ . The total scatter matrix of  $Y_k$  is  $W^T S_T W$ . The optimal transformation  $W_{opt}^1$  is chosen as (O-P in Fig. 1),

$$W_{opt}^1 = \arg \max_W |W^T S_T W| = [W_1^1, W_2^1, \dots, W_m^1]$$

For data representation we have  $\hat{x} = \sum_{i=1}^m y_i v_i$  where  $m < n$ , and  $v_i$  is the eigenvector corresponding to the  $i$ th largest eigenvalue.

In a non-linear transform, replace  $y_i$  by a function of  $y_i, g_i(y_i)$

To obtain maximum class separability according to the linear discriminant criterion, the optimal transformation:  $W_{opt}^2$  is chosen if  $S_w$  is nonsingular. as (O-Z in Fig. 1)

$$W_{opt}^2 = \arg \max_W \frac{|W^T S_b W|}{|W^T S_w W|} = [W_1^2, W_2^2, \dots, W_m^2]$$

It is obvious that  $W_{opt}^1 \neq W_{opt}^2$ .

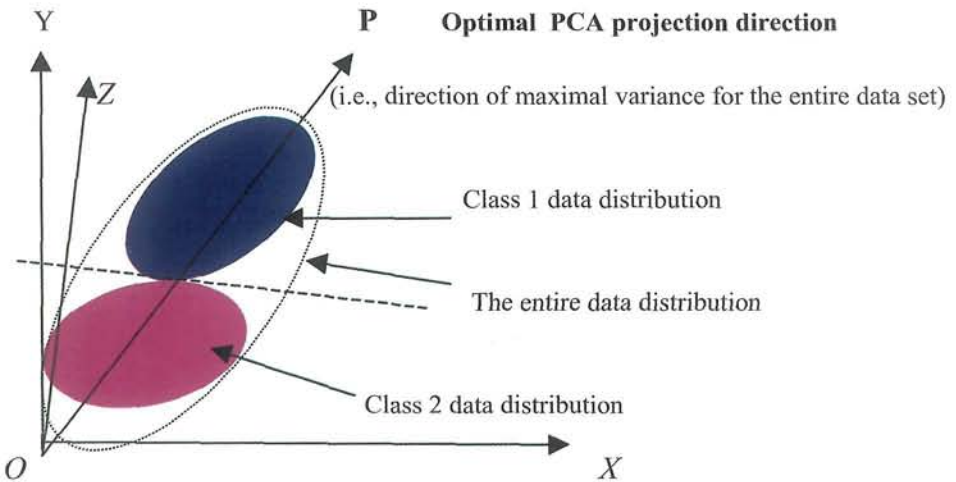


Fig. 1 The optimal PCA projection and separation projection for two sets of data. (O-Z direction is *the least overlapped projection* direction for the two classes)



The linear discriminant analysis (LDA) is generally used in multiple classes also. The following are some drawbacks.

1. The between-class scatter matrix has rank of  $C-1$  where  $C$  is the number of classes. This limits the number of features extracted to  $C-1$ . For remote sensing, the number of classes is usually large.
2. For high-dimensional data, the number of samples per class must also be large, a condition often not satisfied in practice. As a result the within-class scatter matrix can be singular.
3. Existing criteria place too much emphasis on classes which have large inter-class distances.

For discriminant analysis feature extraction, the definition of inter-class scatter can be modified and in fact a weighted inter-class scatter may work well with a larger number of classes and may result in better features without the problem of matrix singularity. This and similar procedures to overcome the drawbacks listed above are now well documented in the pattern recognition literatures (see e.g. [4], [5]).

In summary, PCA or KLT is based on the global covariance matrix. Therefore it is not sensitive to inter-class structure [6]. It is good as a feature reduction tool but the maximum data scatter does not correspond to the best direction for class separability. For discriminant analysis both inter-class and intra-class scatter matrices must be considered. Both EOF and PCA make use of only second order statistics which may not be adequate for good classification. An orthogonal transform that employs high-order statistics is yet to be developed for remote sensing study. Also it is noted that EOF, PCA or KLT involve linear operations which normally do not provide new information from the original data. Functional expansion by other orthogonal basis functions such as Walsh function, Hermite polynomial are possible but they have not been examined for use in remote sensing to the author's knowledge. It is further noted that there are other orthogonal image transforms such as discrete cosine transform, sine transform, slant transform, etc. (see e.g. [7]) but none of which has better minimum mean square reconstruction error than the KLT and the features provided by such transforms are not more effective than the KLT. The discrete cosine transform which has fast transform algorithm available can serve as a useful approximation of KLT which does not have fast computation algorithm. The eigenvectors of PCA or KLT depends on the data. Among other transforms which do not necessarily make use of the orthogonal basis functions are Projection Pursuit [8], Matching Pursuit [9] and mixed transform [10].

### 3. Empirical Orthogonal Functions (EOF)

EOF (empirical orthogonal function) is closely related to PCA and is used in geophysical sciences using a time sequence of images. The characteristics of EOF's make them ideal for spectral compression especially when there are a large number of such bands such as with high resolution images [11-16]. The EOF's form the orthogonal basis in the data space which provides the most efficient data representations. The images corresponding to small eigenvalues are almost entirely due to noise and can be removed to reduce the effect of noises [11]. In general the sequence of images can be considered as a time series of images. In order to obtain a two dimensional data matrix for the analysis, the columns of each image are stacked so that the image becomes a column vector. When these column vector images are placed together as sequential columns, an  $M \times N$  matrix is formed. Here  $M$  represents the

number of elements in the spatial dimension, in this case the number of pixels in the image, and  $N$  represents the number of elements in the temporal dimension. Singular value decomposition (SVD) provides the most efficient and convenient way to calculate EOF's. SVD decomposes the data matrix into three matrices containing the eigenvalues, eigenvectors and eigenfunctions needed for decomposition. In general  $M$  is not equal to  $N$ .

Mathematically, the time series of images is represented by a linear combination of the eigenfunctions  $F_n$ ,

$$T(x,t) = \sum_{n=1}^N a_n(t) F_n(x)$$

where  $a_n$  are the temporal amplitudes, or eigenvectors. The eigenfunctions, or spatial amplitude functions, can themselves be viewed as images, giving a visual representation of the variance of each mode.

Here is another interpretation of EOF. Let  $\Gamma(\Delta)$  be the covariance between  $X(t)$  and  $X(t+\Delta)$ ,  $\Sigma$  be the covariance of  $X(t)$ , and  $\Sigma_{\Delta}$  be the covariance of  $X(t)-X(t+\Delta)$ . The desired linear transform is obtained by solving the generalized eigenvalue problem,

$$\Sigma a = \lambda \Sigma_{\Delta} a \quad \text{where } a \text{ are the (conjugate eigen-) vectors, } \lambda \text{ is the eigenvalue.}$$

#### 4. Component Analysis

The term "component analysis" is in general the subspace analysis in mathematical statistics. In remote sensing it refers to the principal component analysis (PCA), curvilinear component analysis (CCA), canonical component transform (CCT), tasseled cap components (TCT), and independent component analysis (ICA). These component analysis methods are very different conceptually. PCA or KLT is to look for the principal components according to the second order statistics. CCA performs nonlinear feature space transformation while trying to preserve as much as possible the original data information in the lower dimensional space [17]. The nonlinear operation in CCA does present an opportunity for improved features. This is yet to be verified in practice. Both CCT and TCP are well presented in [18]. CCT is similar to PCA, except that the data are not lumped into one distribution in the vector space. Rather, training data for each class are used to maximize the separability of the defined classes. The CCT is a linear transform on the original feature space such that the transformed features are optimized and are arranged in order of decreasing separability of classes. TCT is a linear matrix transform, just as the PCA and CCT, but is fixed and independent of the data. It is, however, sensor dependent and must be derived for any given sensor [18].

ICA looks for independent components from the original data assumed to be linearly mixed from several independent sources. Nonlinear PCA that makes use of the higher order statistical information [19,20] can provide an improvement over the linear PCA that employs only the second order covariance information. Mathematically instead of minimizing the mean square error between  $\mathbf{x}$  and  $\mathbf{W}'\mathbf{y}$ , the nonlinear PCA is to minimize the mean square error between  $\mathbf{x}$  and  $\mathbf{W}'\mathbf{g}(\mathbf{y})$ , where  $\mathbf{g}(\mathbf{y})$  can be an odd function of  $\mathbf{y}$  [19].

ICA is a useful extension of the traditional principle component analysis (PCA). While PCA method attempts to decorrelate the components in a vector, ICA methods are to make the components as independent as possible. There are currently many

approaches available for ICA. Oja et al. developed a nonlinear PCA learning rule [20, 21], Wang et al. developed the bigradient algorithm [22], Hyvarinen developed a fast fixed point algorithm [23], Bell and Sejnowski derived the infomax [24] approach using mutual information and later Lee, et al. developed extended infomax [25] to extract both subgaussian and supergaussian sources. These and many others contributed significantly to the development of ICA all with blind source separation in mind.

ICA applications in remote sensing study become a new topic in recent years. Szu used ICA to reduce boundary error with multiple-label-remote-sensing (MLRS) [26]. Chiang, et al. also used ICA in AVIRIS data analysis [27]. Tu used a noise-adjusted version of fast independent component analysis (NAFICA) for unsupervised signature extraction and separation in hyperspectral images [28]. Yoshida and Omatu showed their work on ICA and BP classifier for LANDSAT TM data [29]. With remote sensing in mind we developed a new independent component analysis (ICA) method that makes use of the higher order statistics. The algorithm is presented in the next section. The work is quite different from that of Cardoso [30]. We name it the joint cumulant ICA (JC-ICA) algorithm [31,32]. It can be implemented efficiently by a neural network. Experimental evidence [31] shows that for the SAR image pixel classification, a small subset of ICA features performs a few percentage point better than the use of original data or PCA as features. The significant component images obtained by ICA have less speckle noise and are more informative. Furthermore for hyperspectral images, ICA can be useful for selecting or reconfiguring spectral bands so that the desired objects in the images may be enhanced [32]. Figs. 2 and 3 show respectively an original AVIRIS image and the enhanced image using the JC-ICA approach. The latter has more desired details. In summary, ICA methods provide speckle reduction in SAR images, better features in pixel classification, and significant data reduction in hyperspectral images.

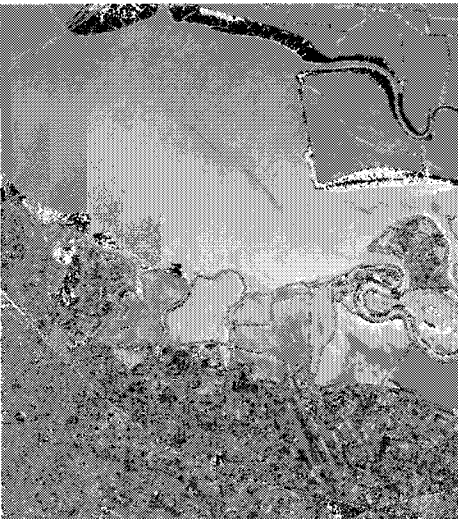


Fig. 2 An AVIRIS image of Moffett field.

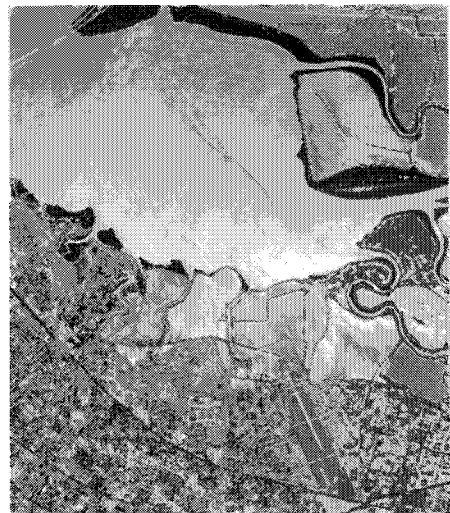


Fig. 3 Enhanced image using JC-ICA.

## 5. The JC-ICA Algorithm

Independence among signals means there is no statistical dependence among them. For the second order statistics with Gaussian random variables, independence means their mutual correlations are zero. For higher order statistics, the dependence is judged by joint cumulants, and independence means that the mutual joint cumulants are zero. Generally speaking, we deal with non-Gaussian random variables. So it is necessary to consider higher order statistics for independence.

Before proceeding with the proposed approach, we would like to review a basic structure of ICA. The structure has 3 levels.  $X$  is the input mixed signal. The first step is to whiten the input data  $X$  and reduce dimension if necessary.  $Y = VX$ .  $V$  is the whitening matrix. Either the standard PCA method or neural learning methods can be used to obtain  $V$ . In our computation,  $V$  is obtained by using eigenvalue-eigenvector method. The purpose of this whitening is to normalize the second order statistics.

The second step is to do separation. Let  $W$  be the separating matrix.  $Z = W^T Y$ .  $Z$  is the extracted independent signals, or at least as independent as possible. There are many algorithms currently available to find  $W$  and  $Z$ . We will discuss ours in this section.

The third step is to estimate the basis vectors of ICA.  $U = QZ$ .  $U$  usually equals  $X$ .  $Q$  is the basis vector, and can be calculated using the neural learning rule:

$$Q(k+1) = Q(k) + \mu(k)(X - Q(k)Z(k))Z(k)^T$$

where  $\mu$  is a small learning rate. This step is useful in projection pursuit. If the goal is merely to do source separation, this step is not necessary.

### 5.1. Initial Approach

Consider a set of  $n$  real random variables  $\{z_1, z_2, \dots, z_n\}$ . From the high order statistics point of view, if these components are statistically independent, their mutual joint cumulants must be zero. In other words zero joint cumulants are a consequence of independence and may not be a sufficient condition for independence. It is reasonable however to expect that when components are as independent as possible, the joint cumulants should be as small as possible. We can determine the independent components by minimizing the joint cumulants of the variables. For computational simplicity, we consider only the 3rd and 4th order cumulants.

We define a penalty function  $P$ ,

$$P(z_1, z_2, \dots, z_j) = \sum_{\text{all } i < j} |\text{Cum}(z_i, z_j, z_i)|/2 + \sum_{\text{all } i < j} |\text{Cum}(z_i, z_j, z_j)|/2 + \sum_{\text{all } i < j} |\text{Cum}(z_i, z_j, z_i, z_j)|$$

The separation matrix  $W$  can be obtained by minimizing  $P$  with regard to  $W$ .  $W$  is  $[w_1^T w_2^T \dots w_n^T]^T$ , where  $w_1$  through  $w_n$  are row vectors. The gradient descent method is used in experiments. Independent components are extracted one by one. When extracting the  $j$ -th component  $x_j$ , joint cumulants are calculated for all  $i$  and  $j$  combination with  $i < j$ .

We first assume input signals  $Y$  be whitened, i.e. zero mean, unit variance. For non-whitened signals, simply do a PCA whitening.

Consider the first component. It can be determined by maximizing its fourth order cumulant, kurtosis [21], and the separating function is,

$$w_1(k+1) = E[Y(Y^T w_1(k))^3] - 3 w_1(k)$$

It is noted that when only one component is considered, there is no joint cumulants. In this case we follow the fixed point approach to extract independent components by maximizing the kurtosis [21]. From the second component, we start using the joint cumulant approach.

For the  $j$ -th component,

$$\begin{aligned} P(j) &= \sum_{i < j} |\text{Cum}(z_i, z_j, z_i)| / 2 + \sum_{i < j} |\text{Cum}(z_i, z_j, z_j)| / 2 + \sum_{i < j} |\text{Cum}(z_i, z_j, z_i, z_j)| \\ &= \sum_{i < j} |E[(w_i Y)^2 \cdot w_j Y]| / 2 + \sum_{i < j} |E[w_i Y \cdot (w_j Y)^2]| / 2 + \\ &\quad \sum_{i < j} |E[(w_i Y)^2 \cdot (w_j Y)^2] - 2 E^2[(w_i Y) \cdot (w_j Y)] - E[(w_i Y)^2] E[(w_j Y)^2]| \end{aligned}$$

and

$$\begin{aligned} \Delta w_j &= \frac{\partial P(j)}{\partial w_j} = \sum_{i < j} \text{Cum}(z_i, z_j, z_i) \frac{\partial \text{Cum}(z_i, z_j, z_i)}{\partial w_j} / 2 + \\ &\quad \sum_{i < j} \text{Cum}(z_i, z_j, z_j) \frac{\partial \text{Cum}(z_i, z_j, z_j)}{\partial w_j} / 2 + \sum_{i < j} \text{Cum}(z_i, z_j, z_i, z_j) \frac{\partial \text{Cum}(z_i, z_j, z_i, z_j)}{\partial w_j} \\ &= \sum_{i < j} E[(w_i Y)^2 \cdot w_j Y'] \cdot E[(w_i Y)^2 Y'] / 2 + \sum_{i < j} E[w_i Y \cdot (w_j Y)^2] \cdot E[(w_i Y) \cdot (w_j Y) Y'] / 2 \\ &\quad + \sum_{i < j} \{E[(w_i Y)^2 \cdot (w_j Y)^2] - 2 E^2[(w_i Y)(w_j Y)] - E[(w_i Y)^2] E[(w_j Y)^2]\} \cdot \\ &\quad \{2E[(w_i Y)^2 \cdot (w_j Y) Y'] - 4E[(w_i Y) \cdot (w_j Y)] E[(w_i Y) Y'] - 2E[(w_i Y)^2] E[(w_j Y) Y']\} \end{aligned}$$

$$w_j(k+1) = w_j(k) + \alpha \Delta w_j.$$

where  $\alpha$  is a small learning rate, and is adjusted adaptively.

To compute the  $j$ -th component,  $j > 1$ , we need to calculate  $3*j - 3$  terms of joint cumulants.

## 5.2. Simplified Approach

We can reduce calculation with simplified joint cumulant combination. The penalty function can be slightly modified to:

$$P(z_1, z_2, \dots, z_k) = \sum_{\substack{i, j < k \\ i, j, k \in s_m}} |\text{Cum}(z_i, z_j, z_k)| + \sum_{\substack{i, j, h < k \\ h, i, j, k \in s_o}} |\text{Cum}(z_i, z_j, z_h, z_k)|$$

To extract the  $k$ -th component, we reduce joint terms by choosing  $i, j, h$  with minimal but essential terms in  $s_m$  and  $s_o$ .  $s_m$  and  $s_o$  are sets of possible combinations for  $z_1, z_2, \dots, z_k$  in the 3rd and 4th joint cumulants.

The initial approach involves a lot of calculation on the joint cumulants and their derivatives. The simplified approach is introduced to reduce the calculation, with a looser condition: we only calculate minimal but essential terms of joint cumulants. All components appear only once in the joint cumulant terms of the same order. Although there is no mathematical proof of this independence criterion, experiments show the results are as good as initial approach, with shorter convergence time.

We still use the gradient descent method to calculate separation matrix  $W$ , and the independent components are again computed one by one. The first two components are computed in the same way as the initial approach because there is no difference in these two cases. For other components, we use the 7th component as an example.

$$C_{71} = \text{Cum}(z_1, z_2, z_7) = E(z_1 z_2 z_7)$$

$$\Delta w_{71} = \frac{\partial \text{Cum}(z_1, z_2, z_7)}{\partial w_7} = E[(w_1 Y \cdot w_2 Y) Y']$$

$$C_{72} = \text{Cum}(z_3, z_4, z_7) = E(z_3 z_4 z_7)$$

$$\Delta w_{72} = \frac{\partial \text{Cum}(z_3, z_4, z_7)}{\partial w_7} = E[(w_3 Y \cdot w_4 Y) Y']$$

$$C_{73} = \text{Cum}(z_5, z_6, z_7) = E(z_5 z_6 z_7)$$

$$\Delta w_{73} = \frac{\partial \text{Cum}(z_5, z_6, z_7)}{\partial w_7} = E[(w_5 Y \cdot w_6 Y) Y']$$

$$C_{74} = \text{Cum}[z_1, z_2, z_3, z_7] = E(z_1 z_2 z_3 z_7) - E(z_1 z_2)E(z_3 z_7) - E(z_1 z_3)E(z_2 z_7) - E(z_1 z_7)E(z_2 z_3)$$

$$\begin{aligned} \Delta w_{74} &= \frac{\partial \text{Cum}(z_1, z_2, z_3, z_7)}{\partial w_7} \\ &= E[(w_1 Y) \cdot (w_2 Y) \cdot (w_3 Y) Y'] - E[(w_1 Y) \cdot (w_2 Y)] E[(w_3 Y) Y'] - \\ &\quad E[(w_1 Y) \cdot (w_3 Y)] E[(w_2 Y) Y'] - E[(w_2 Y) \cdot (w_3 Y)] E[(w_1 Y) Y'] \end{aligned}$$

$$C_{75} = \text{Cum}[z_4, z_5, z_6, z_7] = E(z_4 z_5 z_6 z_7) - E(z_4 z_5)E(z_6 z_7) - E(z_4 z_6)E(z_5 z_7) - E(z_4 z_7)E(z_5 z_6)$$

$$\begin{aligned} \Delta w_{75} &= \frac{\partial \text{Cum}(z_4, z_5, z_6, z_7)}{\partial w_7} \\ &= E[(w_4 Y) \cdot (w_5 Y) \cdot (w_6 Y) Y'] - E[(w_4 Y) \cdot (w_5 Y)] E[(w_6 Y) Y'] - \\ &\quad - E[(w_4 Y) \cdot (w_6 Y)] E[(w_5 Y) Y'] - E[(w_5 Y) \cdot (w_6 Y)] E[(w_4 Y) Y'] \end{aligned}$$

$$\Delta w_7 = C_{71} \Delta w_{71} + C_{72} \Delta w_{72} + C_{73} \Delta w_{73} + C_{74} \Delta w_{74} + C_{75} \Delta w_{75}$$

$$w_7(k+1) = w_7(k) + \alpha \Delta w_7.$$

Other components can be extracted in a similar manner. To extract the  $m$ -th component, there are usually  $\text{ceil}[(m-1)/2] + \text{ceil}[(m-1)/3]$  terms in calculating  $\Delta w$ .  $\text{ceil}()$  is defined as an operation rounding to the nearest larger integer. For the 7th component, it has  $3*7-3 = 18$  terms in initial approach, compared to  $\text{ceil}[(7-1)/2] + \text{ceil}[(7-1)/3] = 5$  terms in simplified approach. Because of the similarity of each term, all derivation and calculation can be done automatically in computer.

The stopping criterion for the algorithm is chosen such that the iteration continues until the penalty function drops below 0.005.

It is important to note that our algorithm is very different from that of Cardoso [30] which uses the likelihood contrast and a class of Jacobi algorithms for optimization. Our method was motivated by the fixed point approach of Oja et al.[21]

## 6. Concluding Remarks

While we are looking for “magic” transforms, either through iterative or non-iterative procedures, that can provide us the desired characteristic, be it effective representation, compression or discrimination, we have to be realistic that such transform may not exist.

Preprocessing and image modeling can be very useful to achieve a better transform. By representing the spatial dependence with a Gaussian Markov random

field, and the PCA dimensionality reduction, the resulting transform can be much more effective for classification [33]. The drawback is that a significant amount of computation is required.

We did not discuss the linear wavelet transform and related transforms since the topic is well examined in chapters in this volume. However, it is worth to note that nonlinear wavelet transforms can also be powerful transforms. Indeed there are many potentials for new and more effective transforms in the foreseeable future for remote sensing, as we constantly seek for better information processing techniques.

## References

1. Richards, J.A. and X. Jia, "*Remote Sensing Digital Image Analysis, An Introduction*", 3<sup>rd</sup> edition, Springer, 1999.
2. Jia, X. and J.A. Richards, "Sequential principal components transformation for efficient hyperspectral remote sensing image display and classification", *IEEE Trans. on Geoscience and Remote Sensing*, vol. 37, pp. 538-542, 1999.
3. Lee, J.S. and K. Hoppel, "Principal components transformation of multifrequency polarimetric SAR imagery", *IEEE Trans. on Geoscience and Remote Sensing*, vol. 29, 1991.
4. Long, M., R.P.W. Duin and R. Haeb-Umbach, "Multiclass linear dimension reduction by weighted pairwise Fisher criteria", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 7, pp. 762-766, July 2001.
5. Bing, Y., J. Liang and C. Ping, "A new LDA-based method for face recognition", *Proc. of the ICPR*, vol. 1, p. 168, August 2002.
6. Benediktsson, J.A., "Statistical and neural network pattern recognition methods for remote sensing applications", Chapter 3.3 of "*Handbook of Pattern Recognition and Computer Vision*", edited by C.H. Chen, et al., World Scientific Publishing, Singapore, 1999.
7. Gonzalez, R.C. and R.C. Woods, "*Digital Image Processing*", 1<sup>st</sup> edition, Prentice-Hall 1992.
8. Jimenez, L.O. and D.A. Landgrebe, "Hyperspectral data analysis and supervised feature reduction with projection pursuit", *IEEE Trans. on Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2653-2667, Nov. 1999.
9. Mallat, S. and Z. Zhang, "Matching pursuits with time-frequency dictionaries", *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3397-3415, Dec. 1993.
10. Alparone, I., F. Argenti and M. Dionisio, "Hyperspectral data analysis by mixed transforms", *Proc. of IGARSS 2002*, Toronto, June 2002.
11. Chen, C.-C.T. and D.A. Landgrebe, "A spectral feature design system for the HIRIS/MODIS era", *IEEE Trans. on Geoscience and Remote Sensing*, vol. 27, pp. 681-686, 1989.
12. Hoffman, R.N. and D.W. Johnson, "Application of EOF's to multispectral imagery: data compression and noise detection for AVIRIS", *IEEE Trans. on Geoscience and Remote Sensing*, vol. 32, pp. 25-34, 1994.
13. Keiner, L.E. and X.H. Yan, "Empirical orthogonal function analysis of sea surface temperature patterns in Delaware Bay", *IEEE Trans. on Geoscience and Remote Sensing*, vol. 35, pp. 1299-1306, 1997.
14. Nielsen, A.A., et al., "A temporal extension to traditional empirical orthogonal function (EOF) analysis", in "*Analysis of Multi-temporal Remote Sensing Images*", edited by L. Bruzzone and P. Smits, World Scientific Publishing (WSP) 2002.

15. K. B. Hilger, et al., "An ACE-based nonlinear extension to traditional EOF analysis", chapter in *"Analysis of Multi-temporal Remote Sensing Images"*, edited by L. Bruzzone and P. Smits, WSP 2002.
16. A.A. Nielsen, et al., "A bivariate extension to traditional EOF analysis", chapter in *"Analysis of Multi-temporal Remote Sensing Images"*, edited by L. Bruzzone and P. Smits, WSP 2002
17. Lennon, M., G. Mercier and M.C. Mouchot, "Curvilinear component analysis for nonlinear dimensionality reduction of hyperspectral images", SPIE Remote Sensing Symposium Conference 4541, Image and Signal Processing for Remote Sensing VII, chaired by S. Serpico, Toulouse, France, Sept. 2001.
18. J.C. Tilton, D. Landgrebe and R.A. Schowengerdt, "Information processing for Remote sensing", chapter in *Encyclopedia of Electrical and Electronics Engineering Series in the area of Geoscience and Remote Sensing*, edited by J.G. Webster, Wiley 1998.
19. Zeng, X., Y.W. Chen and Z. Nakao, "Image feature representation by the subspace of nonlinear PCA", Proc. of the ICPR2002, Quebec City, Aug. 2002.
20. Oja, E., "The nonlinear PCA learning rule in independent component analysis" *Neurocomputing*, vol. 17, no. 1, 1997.
21. Hyvarinen, A., J. Karhunen and E. Oja, "Independent Component Analysis", Wiley 2001.
22. Wang, L. and J. Karhunen, "A unified neural bigradient algorithm for robust PCA and MCA", *Int. J. Neural Syst*, 7, 1996
22. Hyvarinen, A., "A fast fixed-point algorithm for independent component analysis", *Neural Computation*, vol. 9, no. 7, 1997.
23. Bell, A.J. and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution", *Neural Computation*, vol. 7, no. 6, 1995.
24. Lee, T.W., M. Girolami and T. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources", *Neural Computation*, vol. 11, no. 2, pp. 417-441, 1999.
25. Szu, H. and J. Buss, "ICA neural net to refine remote sensing with multiple labels", *Proceedings of SPIE - The International Society for Optical Engineering Wavelet Applications VII Apr 26- 28 2000 v4056*, 32-49, 2000.
27. Chiang, S., C.I. Chang, and I. Ginsberg, "Unsupervised hyperspectral image analysis using independent component analysis", Proc. of IGARSS 2000, Hawaii, July 2000.
28. Tu, T., "Unsupervised signature extraction and separation in hyperspectral images: A noise-adjusted fast independent component analysis approach", *Optical Engineering*, vol. 39, 2000.
29. Yoshida, T. and S. Omatu, "Pattern recognition with neural networks", Proc. of IGARSS 2000, Hawaii, July 2000.
30. Cardoso, J., "High-order contrasts for independent component analysis" *Neural Computation*, vol. 11, pp. 157-192, 1999.
31. Zhang, X. and C.H. Chen, "A new independent component analysis (ICA) method and its application to remote sensing images", *Journal of VLSI and Signal Processing*, to appear.
32. Zhang, X. and C.H. Chen, "On a new independent component analysis (ICA) method using higher order statistics with application to remote sensing image", *Optical Engineering*, July 2002
33. Chen, C.H., and H. Chen, "Unsupervised classification of hyperspectral images using Gauss-Markov random fields and PCA", chapter in this volume.



# **Data Mining**

This page is intentionally left blank

## CHAPTER 3

# SCENE MODELING AND IMAGE MINING WITH A VISUAL GRAMMAR

Selim Aksoy, Carsten Tusk, Krzysztof Koperski, Giovanni Marchisio

*Insightful Corporation*

*1700 Westlake Ave. N., Suite 500, Seattle, WA 98109, USA*

*E-mail: {saksoy,ctusk,krisk,giovanni}@insightful.com*

Automatic content extraction, classification and content-based retrieval are highly desired goals in intelligent remote sensing databases. Pixel level processing has been the common choice for both academic and commercial systems. We extend the modeling of remotely sensed imagery to three levels: Pixel level, region level and scene level. Pixel level features are generated using unsupervised clustering of spectral values, texture features and ancillary data like digital elevation models. Region level features include shape information and statistics of pixel level feature values. Scene level features include statistics and spatial relationships of regions. This chapter describes our work on developing a probabilistic visual grammar to reduce the gap between low-level features and high-level user semantics, and to support complex query scenarios that consist of many regions with different feature characteristics. The visual grammar includes automatic identification of region prototypes and modeling of their spatial relationships. The system learns the prototype regions in an image collection using unsupervised clustering. Spatial relationships are represented by fuzzy membership functions. The system automatically selects significant relationships from training data and builds visual grammar models which can also be updated using user relevance feedback. A Bayesian framework is used to automatically classify scenes based on these models. We demonstrate our system with query scenarios that cannot be expressed by traditional region or scene level approaches but where the visual grammar provides accurate classifications and effective retrieval.

### 1. Introduction

Remotely sensed imagery has become an invaluable tool for scientists, governments, military, and the general public to understand the world and its surrounding environment. Automatic content extraction, classification and content-based retrieval are highly desired goals in intelligent remote sensing databases. Most of the current systems use spectral information or texture features as the input for statistical classifiers that are built using unsupervised or supervised algorithms. The most commonly used classifiers are the minimum distance classifier and the maximum likelihood classifier with a Gaussian density assumption. Spectral signatures and

texture features do not always map conceptually similar patterns to nearby locations in the feature space and limit the success of minimum distance classifiers. Furthermore, these features do not always have Gaussian distributions so maximum likelihood classifiers with this assumption fail to model the data. Image retrieval systems also use spectral or texture features<sup>2</sup> to index images and then apply distance measures<sup>3</sup> in these feature spaces to find similarities. However, there is a large semantic gap between the low-level features and the high-level user expectations and search scenarios.

Pixel level processing has been the common choice for both academic and commercial land cover analysis systems where classifiers have been applied to pixel level measurements. Even though most of the proposed algorithms use pixel level information, remote sensing experts use spatial information to interpret the land cover. Hence, existing systems can only be partial tools for sophisticated analysis of remotely sensed data where a significant amount of expert involvement becomes inevitable. This motivated the research on developing algorithms for region-based analysis with examples including conceptual clustering,<sup>38</sup> region growing<sup>9</sup> and Markov random field models<sup>34</sup> for segmentation of natural scenes; hierarchical segmentation for image mining;<sup>41</sup> rule-based region classification for flood monitoring;<sup>20</sup> region growing for object level change detection;<sup>13</sup> boundary delineation of agricultural fields;<sup>32</sup> and task-specific region merging for road extraction and vegetation area identification.<sup>42</sup>

Traditional region or scene level image analysis algorithms assume that the regions or scenes consist of uniform pixel feature distributions. However, complex query scenarios and image scenes of interest usually contain many pixels and regions that have different feature characteristics. Furthermore, two scenes with similar regions can have very different interpretations if the regions have different spatial arrangements. Even when pixels and regions can be identified correctly, manual interpretation is necessary for studies like landing zone and troop movement planning in military applications and public health and ecological studies in civil applications. Example scenarios include studies on the effects of climate change and human intrusion into previously uninhabited tropical areas, and relationships between vegetation coverage, wetlands and habitats of animals carrying viruses that cause infectious diseases like malaria, West Nile fever, Ebola hemorrhagic fever and tuberculosis.<sup>12,30</sup> Remote sensing imagery with land cover maps and spatial analysis is used for identification of risk factors for locations to which infections are likely to spread. To assist developments in new remote sensing applications, we need a higher level visual grammar to automatically describe and process these scenarios.

Insightful Corporation's VISIMINE system<sup>17,18</sup> supports interactive classification and retrieval of remotely sensed images by modeling them on pixel, region and scene levels. Pixel level characterization provides classification details for each pixel with regard to its spectral, textural and ancillary (e.g. DEM or other GIS layers) attributes. Following a segmentation process that computes an approximate poly-

gon decomposition of each scene, region level features describe properties shared by groups of pixels. Scene level features describe statistical summaries of pixel and region level features, and the spatial relationships of the regions composing a scene. This hierarchical scene modeling bridges the gap between feature extraction and semantic interpretation. VISIMINE also provides an interactive environment for training customized semantic labels from a fusion of visual attributes.

Overviews of different algorithms in VISIMINE were presented in our recent papers.<sup>26,25,23,24,19,17,16,18</sup> This chapter describes our work on developing a probabilistic visual grammar<sup>4</sup> for scene level image mining. Our approach includes learning prototypes of primitive regions and their spatial relationships for higher-level content extraction, and automatic and supervised algorithms for using the visual grammar for content-based retrieval and classification.

Early work on spatial relationships of regions in image retrieval literature included the VisualSEEk project<sup>37</sup> where Smith and Chang used representative colors, centroid locations and minimum bounding rectangles to index regions, and computed similarities between region groups by matching them according to their colors, absolute and relative locations. Berretti *et al.*<sup>5</sup> used four quadrants of the Cartesian coordinate system to compute the directional relationship between a pixel and a region in terms of the number of pixels in the region that were located in each of the four quadrants around that particular pixel. Then, they extended this representation to compute the relationship between two regions using a measure of the number of pairs of pixels in these regions whose displacements fell within each of the four directional relationships. Centroids and minimum bounding rectangles are useful when regions have circular or rectangular shapes but regions in natural scenes often do not follow these assumptions.

Previous work on modeling of spatial relationships in remote sensing applications utilized the concept of spatial association rules. Spatial association rules<sup>15,24</sup> represent topological relationships between spatial objects, spatial orientation and ordering, and distance information. A spatial association rule is of the form  $X \rightarrow Y(c\%)$ , where  $X$  and  $Y$  are sets of spatial or non-spatial predicates and  $c\%$  is the confidence of the rule. An example spatial association rule is *prevalent\_endmember*( $x, concrete$ )  $\wedge$  *texture\_class*( $x, c_1$ )  $\rightarrow$  *close\_to*( $x, coastline$ ) (60%). This rule states that 60% of regions where concrete is the prevalent endmember and that texture features belong to class  $c_1$  are close to a coastline. Examples of spatial predicates include topological relations such as *intersect*, *overlap*, *disjoint*, spatial orientations such as *left\_of* and *west\_of*, and distance information such as *close\_to* or *far\_away*.

Similar work has also been done in the medical imaging area but it usually requires manual delineation of regions by experts. Shyu *et al.*<sup>36</sup> developed a content-based image retrieval system that used features locally computed from manually delineated regions. Neal *et al.*<sup>28</sup> developed topology, part-of and spatial association networks to symbolically model partitive and spatial adjacency relationships of anatomical entities. Tang *et al.*<sup>39,40</sup> divided images into small sub-windows, and

trained neural network classifiers using color and Gabor texture features computed from these sub-windows and the labels assigned to them by experts. These classifiers were then used to assign labels to sub-windows in unknown images, and the labels were verified using a knowledge base of label spatial relationships that was created by experts. Petrakis and Faloutsos<sup>29</sup> used attributed relational graphs to represent features of objects and their relationships in magnetic resonance images. They assumed that the graphs were already known for each image in the database and concentrated on developing fast search algorithms. Chu *et al.*<sup>7</sup> described a knowledge-based semantic image model to represent image objects' characteristics. Graph models are powerful representations but are not usable due to the infeasibility of manual annotation in large databases. Different structures in remote sensing images have different sizes so fixed sized grids cannot capture all structures either.

Our work differs from other approaches in that recognition of regions and decomposition of scenes are done automatically, and training of classifiers requires only a small amount of supervision in terms of example images for classes of interest. The rest of the chapter is organized as follows. An overview of hierarchical scene modeling is given in Sec. 2. The concept of prototype regions is defined in Sec. 3. Spatial relationships of these prototype regions are described in Sec. 4. Algorithms for image retrieval and classification using the spatial relationship models are discussed in Secs. 5 and 6, respectively. Conclusions are given in Sec. 7.

## 2. Hierarchical Scene Modeling

In VISIMINE, we extend the modeling of remotely sensed imagery to three levels: Pixel level, region level and scene level. Pixel level representations include land cover labels for individual pixels (e.g. water, soil, concrete, wetland, conifer, hardwood). Region level representations include shape information and labels for groups of pixels (e.g. city, residential area, forest, lake, tidal flat, field, desert). Scene level representations include interactions of different regions (e.g. forest near a water source, city surrounded by mountains, residential area close to a swamp). This hierarchical scene representation aims to bridge the gap between data and high-level semantic interpretation.

The analysis starts from raw data. Then, features are computed to build classification models for information fusion in terms of structural relationships. Finally, spatial relationships of these basic structures are computed for higher level modeling. Levels of the representation hierarchy are described below.

### 2.1. Raw Data

The lowest level in the hierarchy is the raw data. This includes multispectral data and ancillary data like Digital Elevation Models (DEM) or GIS layers. Examples are given in Figs. 1(a)–1(b).

## 2.2. Features

Feature extraction is used to achieve a higher level of information abstraction and summarization above raw data. To enable processing in pixel, region and scene levels, we use the following state-of-the-art feature extraction methods:

- Pixel level features:
  - (1) Statistics of multispectral values,
  - (2) Spectral unmixing for surface reflectance (spectral mixture analysis),<sup>25</sup>
  - (3) Gabor wavelet features for microtexture analysis,<sup>22</sup>
  - (4) Gray level co-occurrence matrix features for microtexture analysis,<sup>10</sup>
  - (5) Laws features for microtexture analysis,<sup>21</sup>
  - (6) Elevation, slope and aspect computed from DEM data,
  - (7) Unsupervised clustering of spectral or texture values.
- Region level features:
  - (1) Segmentation to find region boundaries (a Bayesian segmentation algorithm under development at Insightful, a hierarchical segmentation algorithm,<sup>41</sup> and a piecewise-polynomial multiscale energy-based region growing segmentation algorithm<sup>14</sup>),
  - (2) Shape information as area, perimeter, centroid, minimum bounding rectangle, orientation of the principal axis, moments, and roughness of boundaries,
  - (3) Statistical summaries (relative percentages) of pixel level features for each region.
- Scene level features:
  - (1) Statistical summaries of pixel and region level features for each scene,
  - (2) Spatial relationships of regions in each scene.

VISIMINE provides a flexible tool where new feature extraction algorithms can be added when new data sources of interest are available. Examples for pixel level features are given in Figs. 1(c)–1(h). Example region segmentation results are given in Figs. 2(a) and 2(c).

## 2.3. Structural Relationships

We use a Bayesian label training algorithm with naive Bayes models<sup>35</sup> to perform fusion of multispectral data, DEM data and the extracted features. The Bayesian framework provides a probabilistic link between low-level image feature attributes and high-level user defined semantic structure labels. The naive Bayes model uses the conditional independence assumption and allows the training of class-conditional probabilities for each attribute. Training for a particular semantic label is done using user labeling of pixels or regions as positive or negative examples for that particular label under training. Then, the probability of a pixel or region belonging to that

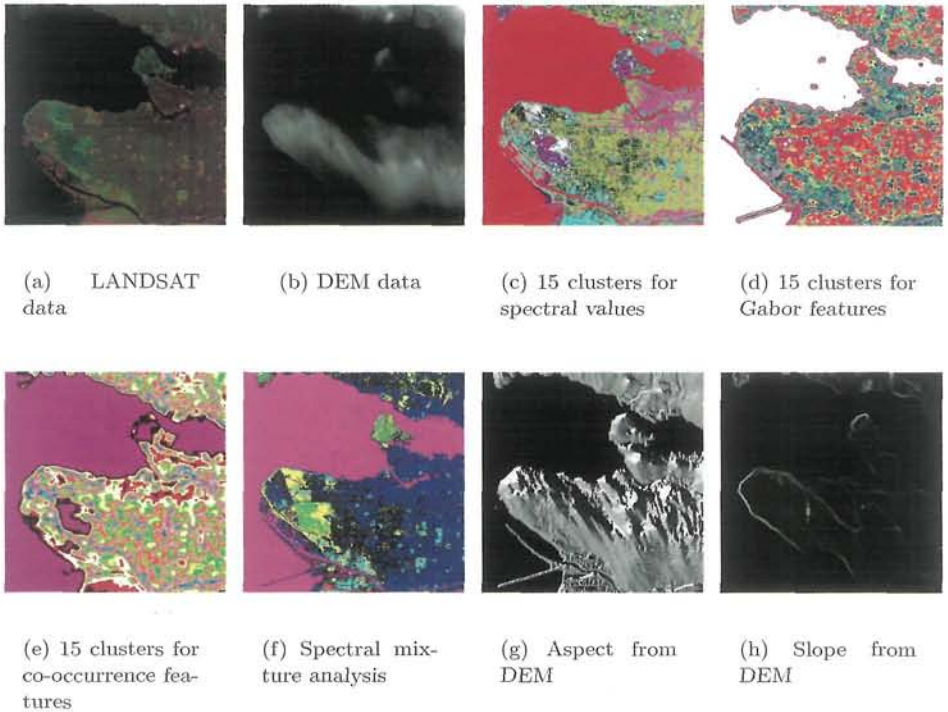


Fig. 1. Raw data and pixel level feature examples for Vancouver, British Columbia. Images in 1(c)-1(f) show the cluster labels for each pixel after unsupervised clustering. Images in 1(g)-1(h) show features computed from DEM data using  $3 \times 3$  windows around each pixel. These pixel level features are used to compute structural relationships for image classification and retrieval.

semantic class is computed as a combination of its attributes using the Bayes rule (e.g. probability of a region being a residential area given its spectral data, texture features and DEM data). Figures 2(b) and 2(d) show examples for labels assigned to regions using the maximum *a posteriori* probability rule.

#### 2.4. Spatial Relationships

The last level in the hierarchy is scene modeling in terms of the spatial relationships of regions. Two scenes with similar regions can have very different interpretations if the regions have different spatial arrangements. Our visual grammar uses region labels identified using supervised and unsupervised classification, and fuzzy modeling of pairwise spatial relationships to describe high-level user concepts (e.g. bordering, invading, surrounding, near, far, right, left, above, below). Fuzzy membership functions for each relationship are constructed based on measurements like region perimeters, shape moments and orientations. When the area of interest consists of multiple regions, the region group is decomposed into region pairs and fuzzy logic



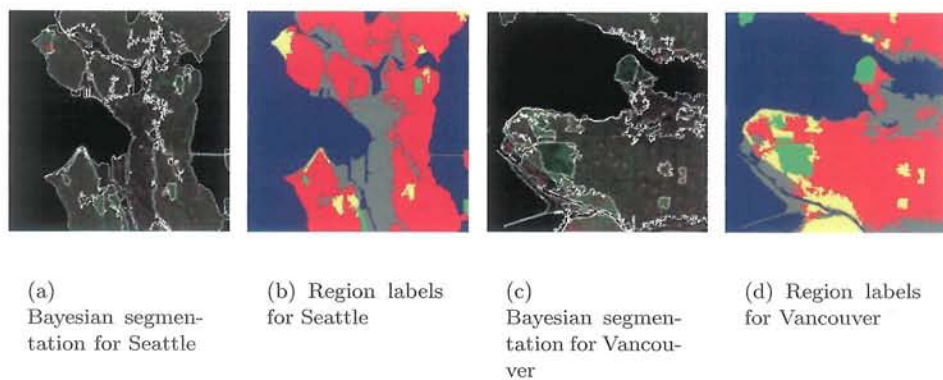


Fig. 2. Region level representation examples for Seattle, Washington and Vancouver, British Columbia. Segmentation boundaries are marked as white. Region labels are city (gray), field (yellow), green park (lime), residential area (red) and water (blue).

is used to combine the measurements on individual pairs.

Combinations of pairwise relationships enable creation of higher level structures that cannot be modeled by individual pixels or regions. For example, an airport consists of buildings, runways and fields around them. An example airport scene and the automatically recognized region labels are shown in Fig. 3. As discussed in Sec. 1, other examples include a landing zone scene which may be modeled in terms of the interactions between flat regions and surrounding hills, public health studies to find residential areas close to swamp areas, and environmental studies to find forests near water sources. The rest of the chapter describes the details of the visual grammar.

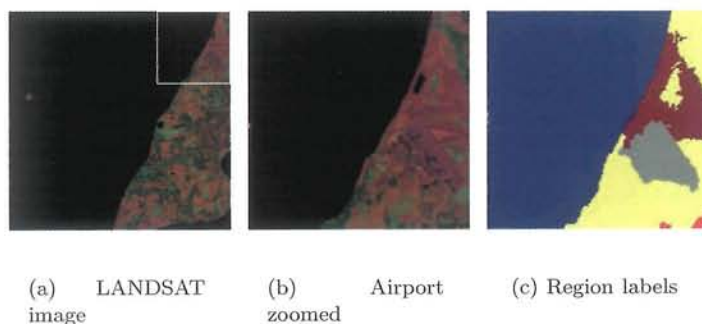


Fig. 3. Modeling of an airport scene in terms of the interactions of its regions. Region labels are dry grass (maroon), buildings and runways (gray), field (yellow), residential area (red), water (blue).

### 3. Prototype Regions

The first step to construct a visual grammar is to find meaningful and representative regions in an image. Automatic extraction of regions is required to handle large amounts of data. To mimic the identification of regions by experts, we define the concept of prototype regions. A prototype region is a region that has a relatively uniform low-level pixel feature distribution and describes a simple scene or part of a scene. Spectral values or any pixel-level feature listed in Sec. 2.2 can be used for region segmentation. Ideally, a prototype is frequently found in a specific class of scenes and differentiates this class of scenes from others. In addition, using prototypes reduces the possible number of associations between regions and makes the combinatorial problem of region matching more tractable. (This will be discussed in detail in Secs. 5 and 6.)

VISI-MINE uses unsupervised  $k$ -means and model-based clustering to automate the process of finding prototypes. Before unsupervised clustering, image segmentation is used to find regions in images. Interesting prototypes in remote sensing images can be cities, rivers, lakes, residential areas, tidal flats, forests, fields, snow, clouds, etc. Figure 4 shows example prototype regions for different LANDSAT images. The following sections describe the algorithms to find prototype regions in an image collection.

#### 3.1. $K$ -means Clustering

$K$ -means clustering<sup>8</sup> is an unsupervised algorithm that partitions the input sample into  $k$  clusters by iteratively minimizing a squared-error criterion function. Clusters are represented by the means of the feature vectors associated with each cluster.

In  $k$ -means clustering the input parameter  $k$  has to be supplied by the user. Once the training data is partitioned into  $k$  groups, the prototypes are represented by the cluster means. Then, Euclidean distance in the feature space is used to match regions to prototypes. The degree of match,  $\tau_{ij}$ , between region  $i$  and prototype  $j$  is computed as

$$\tau_{ij} = \begin{cases} 1 & \text{if } j = \arg \min_{t=1, \dots, k} \|\mathbf{x}_i - \boldsymbol{\mu}_t\|^2 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where  $\mathbf{x}_i$  is the feature vector for region  $i$  and  $\boldsymbol{\mu}_t$  is the mean vector for cluster  $t$ .

#### 3.2. Model-based Clustering

Model-based clustering<sup>8</sup> is also an unsupervised algorithm to partition the input sample. In this case, clusters are represented by parametric density models. Parametric density estimation methods assume a specific form for the density function and the problem reduces to finding the estimates of the parameters of this specific form. However, the assumed form can be quite different from the true density. On

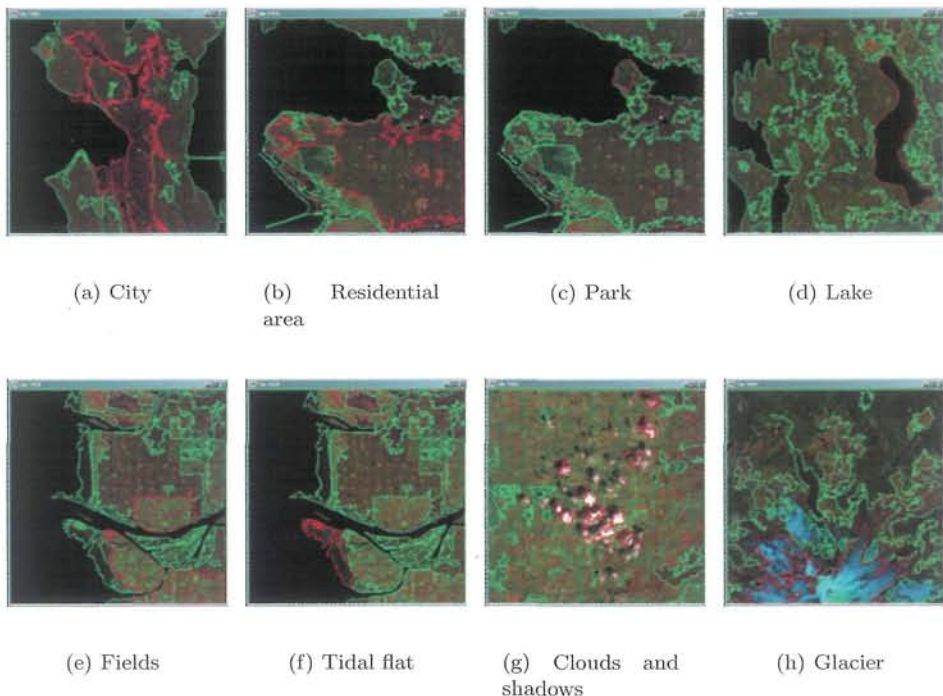


Fig. 4. Example prototype regions for different LANDSAT images. Segmentation boundaries are marked as green and prototype regions are marked as red.

the other hand, non-parametric approaches usually require a large amount of training data and computations can be quite demanding when the data size increases. Mixture models can be used to combine the advantages of both parametric and non-parametric approaches. In a mixture model with  $k$  components, the probability of a feature vector  $\mathbf{x}$  is defined as

$$p(\mathbf{x}) = \sum_{j=1}^k \alpha_j p(\mathbf{x}|j) \quad (2)$$

where  $\alpha_j$  is the mixture weight and  $p(\mathbf{x}|j)$  is the density model for the  $j$ 'th component. Mixture models can be considered as semi-parametric models that are not necessarily restricted to a particular density form but also have a fixed number of parameters independent of the size of the data set.

The most commonly used mixture model is the Gaussian mixture with the component densities defined as

$$p(\mathbf{x}|j) = \frac{1}{(2\pi)^{q/2} |\Sigma_j|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_j)^T \Sigma_j^{-1} (\mathbf{x}-\mu_j)} \quad (3)$$

where  $\mu_j$  is the mean vector and  $\Sigma_j$  is the covariance matrix for the  $j$ 'th com-

ponent respectively, and  $q$  is the dimension of the feature space,  $\mathbf{x} \in \mathbb{R}^q$ . The Expectation-Maximization (EM) algorithm<sup>27</sup> can be used to estimate the parameters of a mixture. The EM algorithm first finds the expected value of the data log-likelihood using the current parameter estimates (expectation step). Then, the algorithm maximizes this expectation (maximization step). These two steps are repeated iteratively. Each iteration is guaranteed to increase the log-likelihood and the algorithm is guaranteed to converge to a local maximum of the likelihood function.<sup>27</sup>

The iterations for the EM algorithm proceed by using the current estimates as the initial estimates for the next iteration. The  $k$ -means algorithm can be used to determine the initial configuration. The mixture weights are computed from the proportion of examples belonging to each cluster. The means are the cluster means. The covariance matrices are calculated as the sample covariance of the points associated with each cluster. Closed form solutions of the EM algorithm for different covariance structures<sup>6,1</sup> are given in Table 1. As a stopping criterion for the EM algorithm, we can use a threshold for the number of iterations or we can stop if the change in log-likelihood between two iterations is less than a threshold.

Table 1. Solutions of the Expectation-Maximization algorithm for a mixture of  $k$  Gaussians.  $\mathbf{x}_1, \dots, \mathbf{x}_n$  are training feature vectors independent and identically distributed with  $p(\mathbf{x})$  as defined in Eq. (2). Covariance structures used are:  $\Sigma_j = \sigma^2 \mathbf{I}$ , all components having the same spherical covariance matrix;  $\Sigma_j = \sigma_j^2 \mathbf{I}$ , each component having an individual spherical covariance matrix;  $\Sigma_j = \text{diag}(\{\sigma_{jt}^2\}_{t=1}^q)$ , each component having an individual diagonal covariance matrix;  $\Sigma_j = \Sigma$ , each component having the same full covariance matrix;  $\hat{\Sigma}_j$ , each component having an individual full covariance matrix.

Variable	Estimate
$p(j \mathbf{x}_i)$	$\frac{\alpha_j p(\mathbf{x}_i j)}{\sum_{k=1}^k \alpha_k p(\mathbf{x}_i k)}$
$\hat{\alpha}_j$	$\frac{\sum_{i=1}^n p(j \mathbf{x}_i)}{\sum_{i=1}^n p(j \mathbf{x}_i)}$
$\hat{\mu}_j$	$\frac{\sum_{i=1}^n p(j \mathbf{x}_i) \mathbf{x}_i}{\sum_{i=1}^n p(j \mathbf{x}_i)}$
$\hat{\Sigma}_j = \hat{\sigma}^2 \mathbf{I}$	$\hat{\sigma}^2 = \frac{\sum_{j=1}^k \sum_{i=1}^n p(j \mathbf{x}_i) (\mathbf{x}_i - \hat{\mu}_j)^T (\mathbf{x}_i - \hat{\mu}_j)}{\sum_{i=1}^n p(j \mathbf{x}_i)}$
$\hat{\Sigma}_j = \hat{\sigma}_j^2 \mathbf{I}$	$\hat{\sigma}_j^2 = \frac{\sum_{i=1}^n p(j \mathbf{x}_i) (\mathbf{x}_i - \hat{\mu}_j)^T (\mathbf{x}_i - \hat{\mu}_j)}{q \sum_{i=1}^n p(j \mathbf{x}_i)}$
$\hat{\Sigma}_j = \text{diag}(\{\hat{\sigma}_{jt}^2\}_{t=1}^q)$	$\hat{\sigma}_{jt}^2 = \frac{\sum_{i=1}^n p(j \mathbf{x}_i) (\mathbf{x}_{it} - \hat{\mu}_{jt})^2}{\sum_{i=1}^n p(j \mathbf{x}_i)}$
$\hat{\Sigma}_j = \hat{\Sigma}$	$\hat{\Sigma} = \frac{\sum_{j=1}^k \sum_{i=1}^n p(j \mathbf{x}_i) (\mathbf{x}_i - \hat{\mu}_j) (\mathbf{x}_i - \hat{\mu}_j)^T}{\sum_{i=1}^n p(j \mathbf{x}_i)}$
$\hat{\Sigma}_j$ , full	$\hat{\Sigma}_j = \frac{\sum_{i=1}^n p(j \mathbf{x}_i) (\mathbf{x}_i - \hat{\mu}_j) (\mathbf{x}_i - \hat{\mu}_j)^T}{\sum_{i=1}^n p(j \mathbf{x}_i)}$

The number of components in the mixture can be either supplied by the user or chosen using optimization criteria like the Minimum Description Length Principle.<sup>31,1</sup> Once the mixture parameters are computed, each component corresponds to a prototype. The degree of match,  $\tau_{ij}$ , between region  $i$  and prototype  $j$  becomes the posterior probability  $\tau_{ij} = p(j|\mathbf{x}_i)$ . The maximum *a posteriori* proba-

bility (MAP) rule is used to match regions to prototypes where region  $i$  is assigned to prototype  $j^*$  as

$$\begin{aligned}
 j^* &= \arg \max_{j=1,\dots,k} p(j|\mathbf{x}_i) \\
 &= \arg \max_{j=1,\dots,k} \alpha_j p(\mathbf{x}_i|j) \\
 &= \arg \max_{j=1,\dots,k} \log(\alpha_j p(\mathbf{x}_i|j)) \\
 &= \arg \max_{j=1,\dots,k} \left\{ \log \alpha_j - \frac{1}{2} \log |\Sigma_j| - \frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_j) \right\}.
 \end{aligned} \tag{4}$$

## 4. Region Relationships

After the regions in the database are clustered into groups of prototype regions, the next step in the construction of the visual grammar is modeling of their spatial relationships. The following sections describe how relationships of region pairs and their combinations can be computed to describe high-level user concepts.

### 4.1. Second-order Region Relationships

Second-order region relationships consist of the relationships between region pairs. These pairs can occur in the image in many possible ways. However, the regions of interest are usually the ones that are close to each other. Representations of spatial relationships depend on the representations of regions. VISIMINE models regions by their boundary pixels and moments. Other possible representations include minimum bounding rectangles,<sup>37</sup> Fourier descriptors<sup>33</sup> and graph-based approaches.<sup>29</sup>

The spatial relationships between all region pairs in an image can be represented by a region relationship matrix. To find the relationship between a pair of regions represented by their boundary pixels and moments, we first compute

- perimeter of the first region,  $\pi_i$
- perimeter of the second region,  $\pi_j$
- common perimeter between two regions,  $\pi_{ij}$
- ratio of the common perimeter to the perimeter of the first region,  $r_{ij} = \frac{\pi_{ij}}{\pi_i}$
- closest distance between the boundary pixels of the first region and the boundary pixels of the second region,  $d_{ij}$
- centroid of the first region,  $\nu_i$
- centroid of the second region,  $\nu_j$
- angle between the horizontal (column) axis and the line joining the centroids,  $\theta_{ij}$

where  $i, j \in \{1, \dots, n\}$  and  $n$  is the number of regions in the image.

The distance  $d_{ij}$  is computed using the distance transform.<sup>11</sup> Given a particular region  $\mathcal{A}$ , to each pixel that is not in  $\mathcal{A}$ , the distance transform assigns a number that is the spatial distance between that pixel and  $\mathcal{A}$ . Then, the distance between

region  $\mathcal{A}$  and another region  $\mathcal{B}$  is the smallest distance transform value for the boundary pixels of  $\mathcal{B}$ . The angle  $\theta_{ij}$  is computed as

$$\theta_{ij} = \begin{cases} \arccos\left(\frac{\nu_{ic} - \nu_{jc}}{d_{ij}}\right) & \text{if } \nu_{ir} \geq \nu_{jr} \\ -\arccos\left(\frac{\nu_{ic} - \nu_{jc}}{d_{ij}}\right) & \text{otherwise} \end{cases} \quad (5)$$

where  $\nu_{ir}$  and  $\nu_{ic}$  are the row and column coordinates of the centroid of region  $i$ , respectively (see Fig. 5 for illustrations). Then, the  $n \times n$  region relationship matrix is defined as

$$\mathbf{R} = \{\{r_{ij}, d_{ij}, \theta_{ij}\} \mid i, j = 1, \dots, n, \forall i \neq j\}. \quad (6)$$

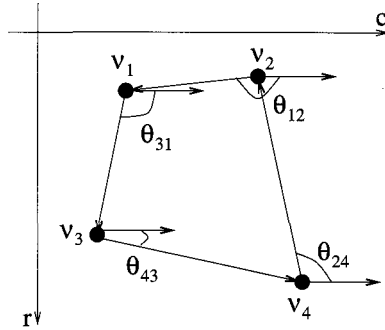
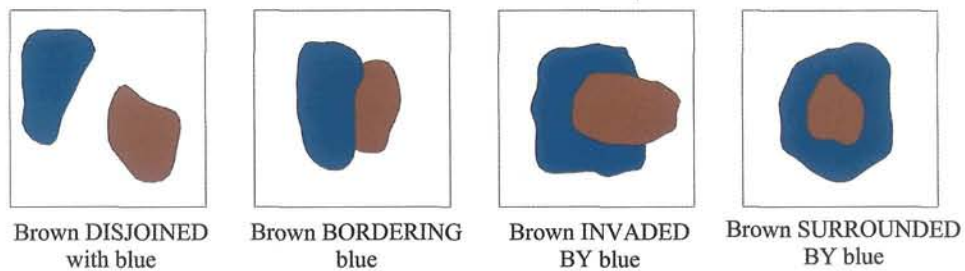


Fig. 5. Orientation of two regions is computed using the angle between the horizontal (column) axis and the line joining their centroids. In the examples above,  $\theta_{ij}$  is the angle between the  $c$ -axis and the line directed from the second centroid  $\nu_j$  to the first centroid  $\nu_i$ . It is used to compute the orientation of region  $i$  with respect to region  $j$ .  $\theta_{ij}$  increases in the clockwise direction, in this case  $\theta_{24} < 0 < \theta_{43} < \theta_{31} < \theta_{12}$ .

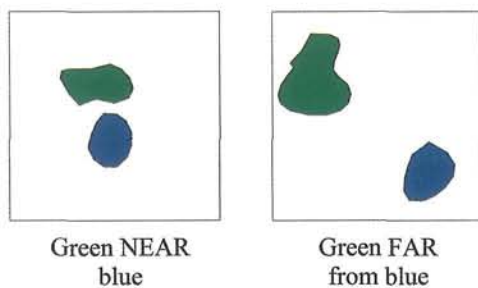
One way to define the spatial relationships between regions  $i$  and  $j$  is to use crisp (Boolean) decisions about  $r_{ij}$ ,  $d_{ij}$  and  $\theta_{ij}$ . Another way is to define them as relationship classes.<sup>33</sup> Each region pair can be assigned a degree of their spatial relationship using fuzzy class membership functions. Denote the class membership functions by  $\Omega_c$  with  $c \in \{\text{DIS, BOR, INV, SUR, NEAR, FAR, RIGHT, LEFT, ABOVE, BELOW}\}$  corresponding to *disjoined*, *bordering*, *invaded\_by*, *surrounded\_by*, *near*, *far*, *right*, *left*, *above* and *below*, respectively. Then, the value  $\Omega_c(r_{ij}, d_{ij}, \theta_{ij})$  represents the degree of membership of regions  $i$  and  $j$  to class  $c$ .

Among the above, *disjoined*, *bordering*, *invaded\_by* and *surrounded\_by* are perimeter-class relationships, *near* and *far* are distance-class relationships, and *right*, *left*, *above* and *below* are orientation-class relationships. These relationships are divided into sub-groups because multiple relationships can be used to describe a region pair, e.g. *invaded\_by* from *left*, *bordering* from *above*, and *near* and *right*, etc. Illustrations are given in Fig. 6.

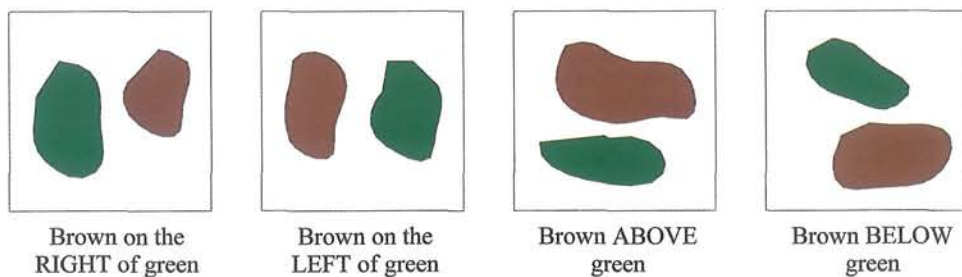
For the perimeter-class relationships, we use the perimeter ratios  $r_{ij}$  with the following trapezoid membership functions:



(a) Perimeter-class relationships: *disjoined*, *bordering*, *invaded\_by* and *surrounded\_by*



(b) Distance-class relationships: *near* and *far*



(c) Orientation-class relationships: *right*, *left*, *above* and *below*

Fig. 6. Spatial relationships of region pairs.

- *disjoined*:

$$\Omega_{\text{DIS}}(r_{ij}) \triangleq \begin{cases} 1 & \text{if } r_{ij} = 0 \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

- *bordering*:

$$\Omega_{\text{BOR}}(r_{ij}) \triangleq \begin{cases} 1 & \text{if } 0 < r_{ij} \leq 0.40 \\ -\frac{20}{13}r_{ij} + \frac{21}{13} & \text{if } 0.40 < r_{ij} \leq 1 \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

- *invaded\_by*:

$$\Omega_{\text{INV}}(r_{ij}) \triangleq \begin{cases} 10r_{ij} - 4 & \text{if } 0.40 \leq r_{ij} < 0.50 \\ 1 & \text{if } 0.50 \leq r_{ij} \leq 0.80 \\ -\frac{10}{3}r_{ij} + \frac{11}{3} & \text{if } 0.80 < r_{ij} \leq 1 \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

- *surrounded\_by*:

$$\Omega_{\text{SUR}}(r_{ij}) \triangleq \begin{cases} \frac{20}{3}r_{ij} - \frac{16}{3} & \text{if } 0.80 \leq r_{ij} < 0.95 \\ 1 & \text{if } 0.95 \leq r_{ij} \leq 1 \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

These functions are shown in Fig. 7(a). The motivation for the choice of these functions is as follows. Two regions are disjoint when they are not touching each other. They are bordering each other when they have a common perimeter. When the common perimeter between two regions gets closer to 50%, the larger region starts invading the smaller one. When the common perimeter goes above 80%, the relationship is considered an almost complete invasion, i.e. surrounding.

For the distance-class relationships, we use the perimeter ratios  $r_{ij}$ , distances between region boundaries  $d_{ij}$  and sigmoid membership functions with the constraint  $\Omega_{\text{NEAR}}(r_{ij}, d_{ij}) + \Omega_{\text{FAR}}(r_{ij}, d_{ij}) = 1$ . The membership functions are defined as:

- *near*:

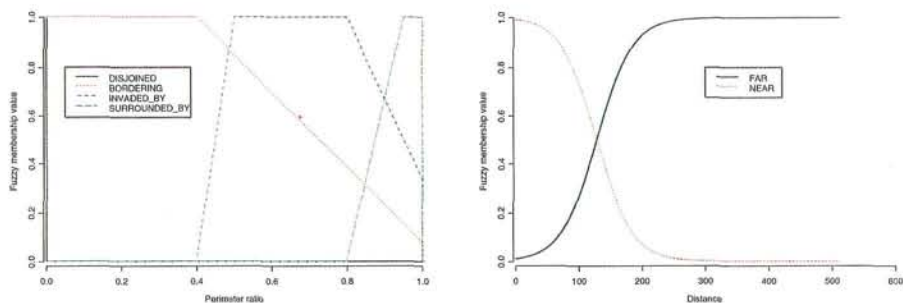
$$\Omega_{\text{NEAR}}(r_{ij}, d_{ij}) \triangleq \begin{cases} 1 & \text{if } r_{ij} > 0 \\ \frac{e^{-\alpha(d_{ij}-\beta)}}{1+e^{-\alpha(d_{ij}-\beta)}} & \text{otherwise.} \end{cases} \quad (11)$$

- *far*:

$$\Omega_{\text{FAR}}(r_{ij}, d_{ij}) \triangleq \begin{cases} 0 & \text{if } r_{ij} > 0 \\ \frac{1}{1+e^{-\alpha(d_{ij}-\beta)}} & \text{otherwise.} \end{cases} \quad (12)$$

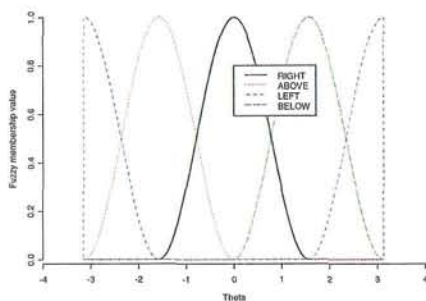
These functions are shown in Fig. 7(b).  $\beta$  is the parameter that determines the cut-off value when a region becomes more far than near, and  $\alpha$  is the parameter that determines the crispness of the function. We first choose  $\beta$  to be a quarter of the image width, i.e.  $\beta = 0.25w$  where  $w$  is the image width, and then choose  $\alpha$  to give a *far* fuzzy membership value less than 0.01 at distance 0, i.e.  $\frac{1}{1+e^{\alpha\beta}} < 0.01 \Rightarrow \alpha > \log(99)/\beta$ .





(a) Perimeter-class spatial relationships

(b) Distance-class spatial relationships



(c) Orientation-class spatial relationships

Fig. 7. Fuzzy membership functions for pairwise spatial relationships.

For the orientation-class relationships, we use the angles  $\theta_{ij}$  and truncated cosine membership functions with the constraint  $\Omega_{\text{RIGHT}}(\theta_{ij}) + \Omega_{\text{LEFT}}(\theta_{ij}) + \Omega_{\text{ABOVE}}(\theta_{ij}) + \Omega_{\text{BELOW}}(\theta_{ij}) = 1$ . The membership functions are defined as:

- *right*:

$$\Omega_{\text{RIGHT}}(\theta_{ij}) \triangleq \begin{cases} \frac{1+\cos(2\theta_{ij})}{2} & \text{if } -\pi/2 < \theta_{ij} < \pi/2 \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

- *left*:

$$\Omega_{\text{LEFT}}(\theta_{ij}) \triangleq \begin{cases} \frac{1+\cos(2\theta_{ij})}{2} & \text{if } -\pi < \theta_{ij} < -\pi/2 \text{ or } \pi/2 < \theta_{ij} < \pi \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

- *above*:

$$\Omega_{\text{ABOVE}}(\theta_{ij}) \triangleq \begin{cases} \frac{1-\cos(2\theta_{ij})}{2} & \text{if } -\pi < \theta_{ij} < 0 \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

- *below*:

$$\Omega_{\text{BELOW}}(\theta_{ij}) \triangleq \begin{cases} \frac{1-\cos(2\theta_{ij})}{2} & \text{if } 0 < \theta_{ij} < \pi \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

These functions are shown in Fig. 7(c).

Note that the pairwise relationships are not always symmetric, i.e.  $\Omega_c(r_{ij}, d_{ij}, \theta_{ij})$  is not necessarily equal to  $\Omega_c(r_{ji}, d_{ji}, \theta_{ji})$ . Furthermore, some relationships are stronger than others. For example, *surrounded\_by* is stronger than *invaded\_by*, and *invaded\_by* is stronger than *bordering*, e.g. the relationship “small region *invaded\_by* large region” is preferred over the relationship “large region *bordering* small region”. The class membership functions are chosen so that only one of them is the largest for a given set of measurements  $r_{ij}, d_{ij}, \theta_{ij}$ . We label a region pair as having the perimeter-class, distance-class and orientation-class relationships

$$\begin{aligned} c_{ij}^1 &= \arg \max_{c \in \{\text{DIS, BOR, INV, SUR}\}} \Omega_c(r_{ij}, d_{ij}, \theta_{ij}) \\ c_{ij}^2 &= \arg \max_{c \in \{\text{NEAR, FAR}\}} \Omega_c(r_{ij}, d_{ij}, \theta_{ij}) \\ c_{ij}^3 &= \arg \max_{c \in \{\text{RIGHT, LEFT, ABOVE, BELOW}\}} \Omega_c(r_{ij}, d_{ij}, \theta_{ij}) \end{aligned} \quad (17)$$

with the corresponding degrees

$$\rho_{ij}^t = \Omega_{c_{ij}^t}(r_{ij}, d_{ij}, \theta_{ij}), \quad t = 1, 2, 3. \quad (18)$$

## 4.2. Higher-order Region Relationships

Higher-order relationships (of region groups) can be decomposed into multiple second-order relationships (of region pairs). Therefore, the measures defined in the previous section can be computed for each of the pairwise relationships and can be combined to measure the combined relationship. The equivalent of the Boolean “and” operation in fuzzy logic is the “min” operation. For a combination of  $k$  regions, there are  $\binom{k}{2} = \frac{k(k-1)}{2}$  pairwise relationships. Therefore, the relationship between these  $k$  regions can be represented as lists of  $\binom{k}{2}$  pairwise relationships using Eq. (17) as

$$c_{1\dots k}^t = \{c_{ij}^t \mid i, j = 1, \dots, k, \forall i < j\}, \quad t = 1, 2, 3 \quad (19)$$

with the corresponding degrees computed using Eq. (18) as

$$\rho_{1\dots k}^t = \min_{\substack{i, j=1, \dots, k \\ i < j}} \rho_{ij}^t, \quad t = 1, 2, 3. \quad (20)$$

Example decompositions are given in Fig. 8. These examples show scenarios that cannot be described by conventional region or scene level image analysis algorithms which assume the regions or scenes consist of pixels with similar feature characteristics.

## 5. Image Retrieval

To use the automatically built visual grammar models for image mining, users can compose queries for complex scene scenarios by giving a set of example regions or by selecting an area of interest in a scene. VISIMINE encodes and searches for a query scene with multiple regions using the visual grammar as follows:

- (1) Let  $k$  be the number of regions selected by the user. Find the prototype label for each of the  $k$  regions.
- (2) Find the perimeter ratio, distance and orientation for each of the  $\binom{k}{2}$  possible region pairs.
- (3) Find the spatial relationship and its degree for these  $k$  regions using Eqs. (19) and (20). Denote them by  $\bar{c}^t = \{\bar{c}_{ij}^t \mid i, j = 1, \dots, k, \forall i < j\}$ ,  $t = 1, 2, 3$  and  $\bar{\rho}^t$ ,  $t = 1, 2, 3$ , respectively.
- (4) For each image in the database,
  - (a) For each query region, find the list of regions with the same prototype label as itself. Denote these lists by  $U_i$ ,  $i = 1, \dots, k$ . These regions are the candidate matches to query regions. Using previously defined prototype labels simplifies region matching into a table look-up process instead of expensive similarity computations between region features.
  - (b) Rank region groups  $(u_1, u_2, \dots, u_k) \in U_1 \times U_2 \times \dots \times U_k$  according to the distance

$$\left| \min_{t=1,2,3} \bar{\rho}^t - \min_{t=1,2,3} \min_{\substack{i,j=1,\dots,k \\ i < j}} \Omega_{\bar{c}_{ij}^t}(r_{u_i u_j}, d_{u_i u_j}, \theta_{u_i u_j}) \right| \quad (21)$$

or alternatively according to

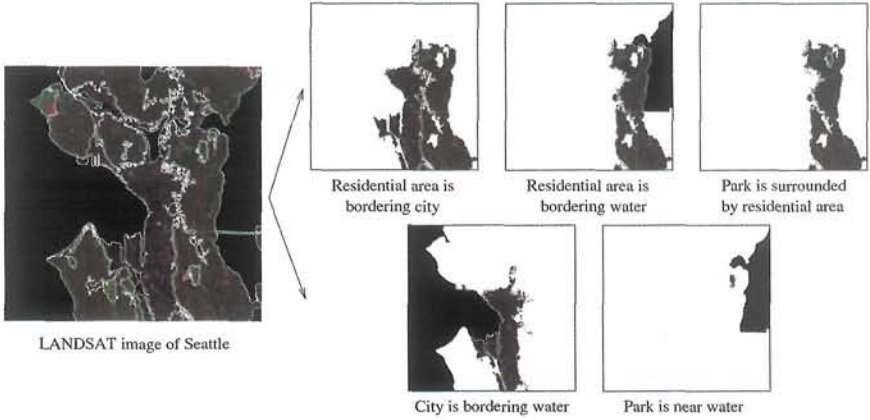
$$\max_{t=1,2,3} \max_{\substack{i,j=1,\dots,k \\ i < j}} \left| \bar{\rho}_{ij}^t - \Omega_{\bar{c}_{ij}^t}(r_{u_i u_j}, d_{u_i u_j}, \theta_{u_i u_j}) \right|. \quad (22)$$

- (c) The equivalent of the Boolean “or” operation in fuzzy logic is the “max” operation. To rank image tiles, use the distance

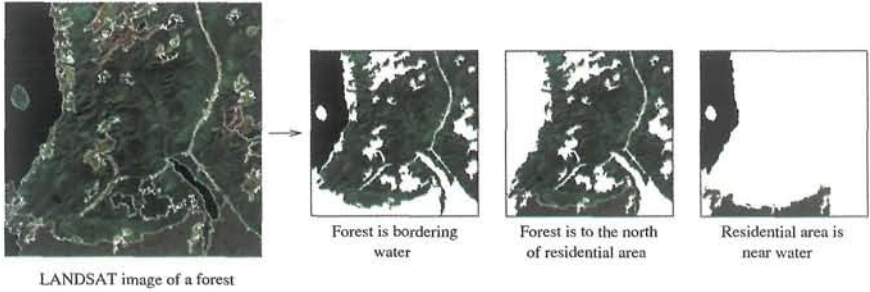
$$\left| \min_{t=1,2,3} \bar{\rho}^t - \max_{\substack{(u_1, u_2, \dots, u_k) \in \\ U_1 \times U_2 \times \dots \times U_k}} \left\{ \min_{t=1,2,3} \min_{\substack{i,j=1,\dots,k \\ i < j}} \Omega_{\bar{c}_{ij}^t}(r_{u_i u_j}, d_{u_i u_j}, \theta_{u_i u_j}) \right\} \right| \quad (23)$$

or alternatively the distance

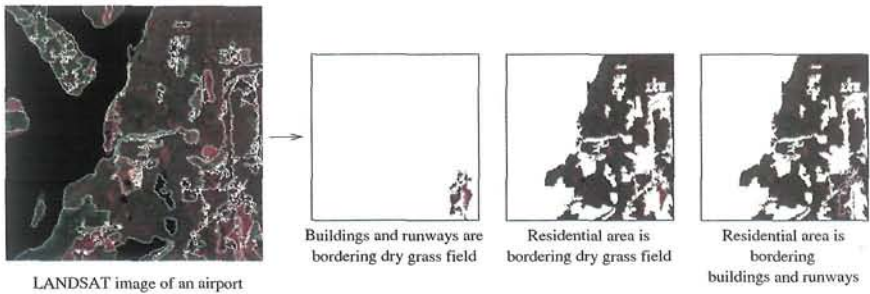
$$\min_{\substack{(u_1, u_2, \dots, u_k) \in \\ U_1 \times U_2 \times \dots \times U_k}} \left\{ \max_{t=1,2,3} \max_{\substack{i,j=1,\dots,k \\ i < j}} \left| \bar{\rho}_{ij}^t - \Omega_{\bar{c}_{ij}^t}(r_{u_i u_j}, d_{u_i u_j}, \theta_{u_i u_j}) \right| \right\}. \quad (24)$$



(a) Relationships among residential area, city, park and water in a Seattle scene



(b) Relationships among forest, water and residential area in a forest scene



(c) Relationships among buildings, runways, dry grass field and residential area in an airport scene

Fig. 8. Example decomposition of scenes into relationships of region pairs. Segmentation boundaries are marked as white.

In some cases, some of the spatial relationships (e.g. *above*, *right*) can be too restrictive. The visual grammar also includes a DONT\_CARE relationship class that allows the user to constrain the searches based on the relationship groups he is interested in using the VISIMINE graphical user interface. Relevance feedback can also be used to find the most important relationship class (perimeter, distance or orientation) for a particular query.

Example queries on a LANDSAT database covering Washington State in the U.S.A. and southern part of British Columbia in Canada are given in Figs. 9–13. Traditionally, queries that consist of multiple regions are handled by computing a single set of features using all the pixels in the union of those regions. However, this averaging causes a significant information loss because features of pixels in different regions usually correspond to different neighborhoods in the feature space and averaging distorts the multimodal characteristic of the query. For example, averaging features computed from the regions in these query scenes ignores the spatial organization of concrete, soil, grass, trees and water in those scenes. On the other hand, the visual grammar can capture both feature and spatial characteristics of region groups.

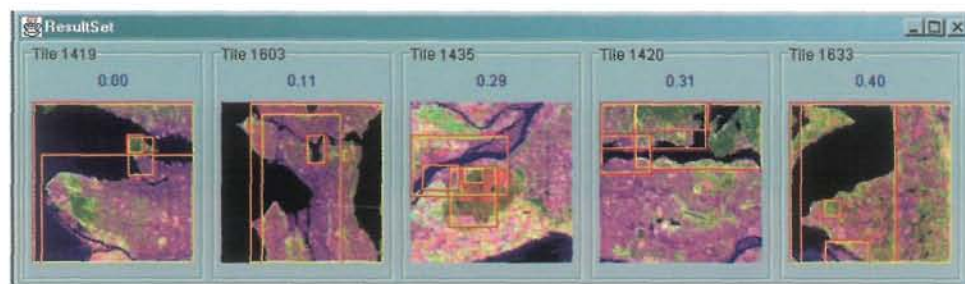


Fig. 9. Search results for a scene where a residential area is bordering a city and both are bordering water, and a park is surrounded by a residential area and is also near water. Identified regions are marked by their minimum bounding rectangles. Decomposition of the query scene is given in Fig. 8(a).

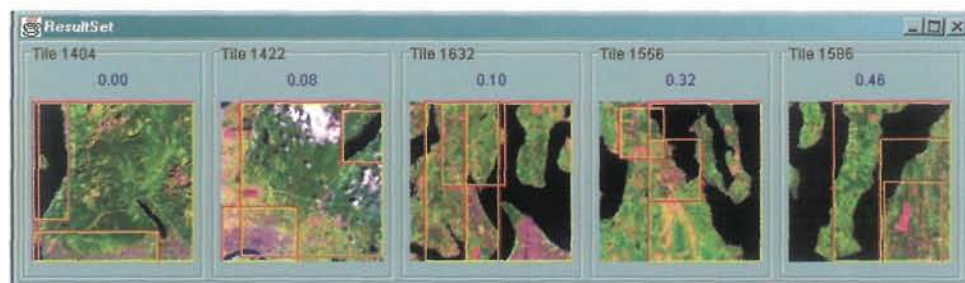


Fig. 10. Search results for a scene where a forest is bordering water and is also to the north of a residential area. Decomposition of the query scene is given in Fig. 8(b).





Fig. 11. Search results for a scene where buildings, runways and their neighboring dry grass field are near a residential area. Decomposition of the query scene is given in Fig. 8(c).



Fig. 12. Search results for a scene where a lake is surrounded by tree covered hills.

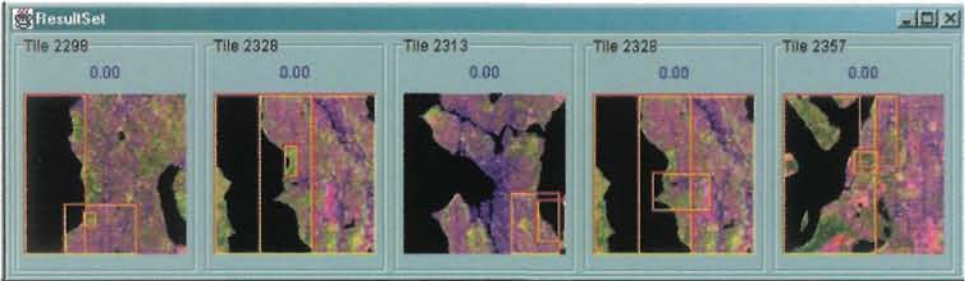


Fig. 13. Search results for a scene where a residential area and its neighboring park are both bordering water.

## 6. Image Classification

Image classification is defined here as a problem of assigning images to different classes according to the scenes they contain. Commonly used statistical classifiers require a lot of training data to effectively compute the spectral and textural signatures for pixels and also cannot do classification based on high-level user concepts because of the lack of spatial information. Rule-based classifiers also require significant amount of user involvement every time a new class is introduced to the

system.

The visual grammar enables creation of higher level classes that cannot be modeled by individual pixels or regions. Furthermore, learning of these classifiers require only a few training images. We use a Bayesian framework that learns scene classes based on automatic selection of distinguishing (e.g. frequently occurring, rarely occurring) relations between regions.

The input to the system is a set of training images that contain example scenes for each class defined by the user. Let  $s$  be the number of classes,  $m$  be the number of relationships defined for region pairs,  $k$  be the number of regions in a region group, and  $t$  be a threshold for the number of region groups that will be used in the classifier. Denote the classes by  $w_1, \dots, w_s$ . VISIMINE automatically builds classifiers from the training data as follows:

- (1) Count the number of times each possible region group with a particular spatial relationship is found in the set of training images for each class. This is a combinatorial problem because the total number of region groups (unordered arrangements without replacement) in an image with  $n$  regions is  $\binom{n}{k}$  and the total number of possible relationships (unordered arrangements with replacement) in a region group is  $\binom{m+\binom{k}{2}-1}{\binom{k}{2}}$ . A region group of interest is the one that is frequently found in a particular class of scenes but rarely exists in other classes. For each region group, this can be measured using class separability which can be computed in terms of within-class and between-class variances of the counts as

$$\varsigma = \log \left( 1 + \frac{\sigma_B^2}{\sigma_W^2} \right) \quad (25)$$

where  $\sigma_W^2 = \sum_{i=1}^s v_i \text{var}\{z_j | j \in w_i\}$  is the within-class variance,  $v_i$  is the number of training images for class  $w_i$ ,  $z_j$  is the number of times this region group is found in training image  $j$ ,  $\sigma_B^2 = \text{var}\{\sum_{j \in w_i} z_j | i = 1, \dots, s\}$  is the between-class variance, and  $\text{var}\{\cdot\}$  denotes the variance of a sample.

- (2) Select the top  $t$  region groups with the largest class separability values. Let  $x_1, \dots, x_t$  be Bernoulli random variables for these region groups, where  $x_j = T$  if the region group  $x_j$  is found in an image and  $x_j = F$  otherwise. Let  $p(x_j = T) = \theta_j$ . Then, the number of times  $x_j$  is found in images from class  $w_i$  has a Binomial( $v_i, \theta_j$ ) =  $\binom{v_i}{v_{ij}} \theta_j^{v_{ij}} (1 - \theta_j)^{v_i - v_{ij}}$  distribution where  $v_{ij}$  is the number of training images for  $w_i$  that contain  $x_j$ . The maximum likelihood estimate of  $\theta_j$  becomes

$$p(x_j = T | w_i) = \frac{v_{ij}}{v_i}. \quad (26)$$

Using a Beta(1,1) distribution as the conjugate prior, the Bayes estimate for  $\theta_j$  is computed as

$$p(x_j = T | w_i) = \frac{v_{ij} + 1}{v_i + 2}. \quad (27)$$

Using a similar procedure with Multinomial and Dirichlet distributions, the Bayes estimate for an image belonging to class  $w_i$  (i.e. containing the scene defined by class  $w_i$ ) is computed as

$$p(w_i) = \frac{v_i + 1}{\sum_{i=1}^s v_i + s}. \quad (28)$$

In other words, discrete probability tables are constructed using  $v_i$  and  $v_{ij}$ ,  $i = 1, \dots, s, j = 1, \dots, t$ , and conjugate priors are used to update them when new images become available via relevance feedback.

- (3) For an unknown image, search for each of the  $t$  region groups (determine whether  $x_j = T$  or  $x_j = F$ ,  $\forall j$ ) and compute the probability for each class using the conditional independence assumption as

$$\begin{aligned} p(w_i|x_1, \dots, x_t) &= \frac{p(w_i, x_1, \dots, x_t)}{p(x_1, \dots, x_t)} \\ &= \frac{p(w_i)p(x_1, \dots, x_t|w_i)}{p(x_1, \dots, x_t)} \\ &= \frac{p(w_i) \prod_{j=1}^t p(x_j|w_i)}{p(x_1, \dots, x_t)}. \end{aligned} \quad (29)$$

Assign that image to the best matching class using the MAP rule as

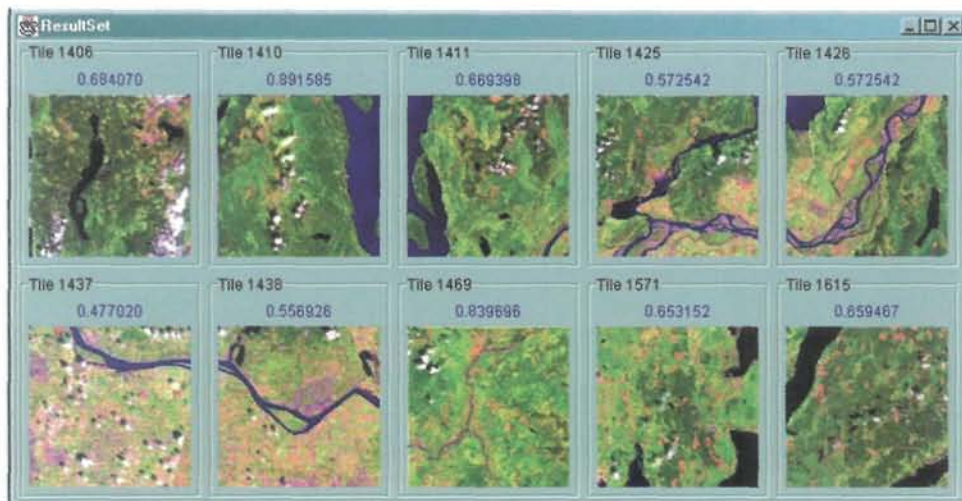
$$\begin{aligned} w^* &= \arg \max_{w_i} p(w_i|x_1, \dots, x_t) \\ &= \arg \max_{w_i} p(w_i) \prod_{j=1}^t p(x_j|w_i). \end{aligned} \quad (30)$$

Classification examples are given in Figs. 14–16. We used four training images for each of the classes defined as “clouds”, “tree covered islands”, “residential areas with a coastline”, “snow covered mountains”, “fields” and “high-altitude forests”. These classes provide a challenge where a mixture of spectral, textural, elevation and spatial information is required for correct identification of the scenes. For example, pixel level classifiers often misclassify clouds as snow and shadows as water. On the other hand, the Bayesian classifier described above could successfully eliminate most of the false alarms by first recognizing regions that belonged to cloud and shadow prototypes and then verified these region groups according to the fact that clouds are often accompanied by their shadows in a LANDSAT scene. Other scene classes like residential areas with a coastline or tree covered islands cannot be identified by pixel level or scene level algorithms that do not use spatial information. The visual grammar classifiers automatically learned the distinguishing region groups that were frequently found in particular classes of scenes but rarely existed in other classes.





(a) Training images

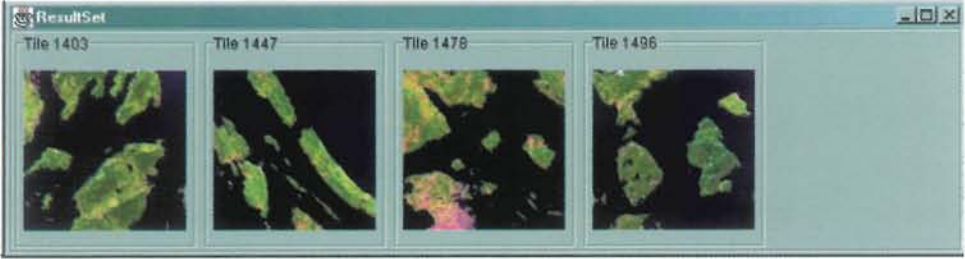


(b) Images classified as containing clouds

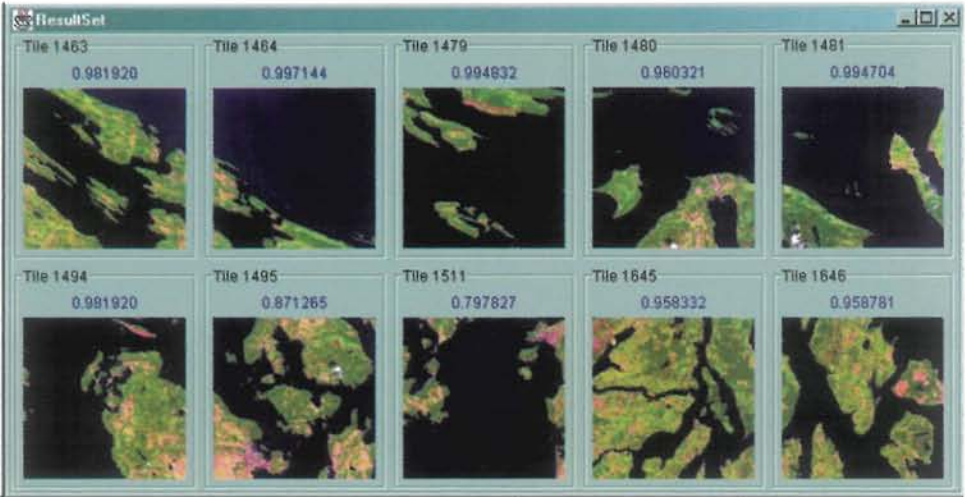
Fig. 14. Classification results for the “clouds” class which is automatically modeled by the distinguishing relationships of white regions (clouds) with their neighboring dark regions (shadows).

## 7. Conclusions

In this chapter we described a probabilistic visual grammar to automatically analyze complex query scenarios using spatial relationships of regions and described algorithms to use it for content-based image retrieval and classification. Our hierarchical scene modeling bridges the gap between feature extraction and semantic interpretation. The approach includes unsupervised clustering to identify prototype regions in images (e.g. city, residential, water, field, forest, glacier), fuzzy modeling of region spatial relationships to describe high-level user concepts (e.g. bordering, surrounding, near, far, above, below), and Bayesian classifiers to learn image classes



(a) Training images

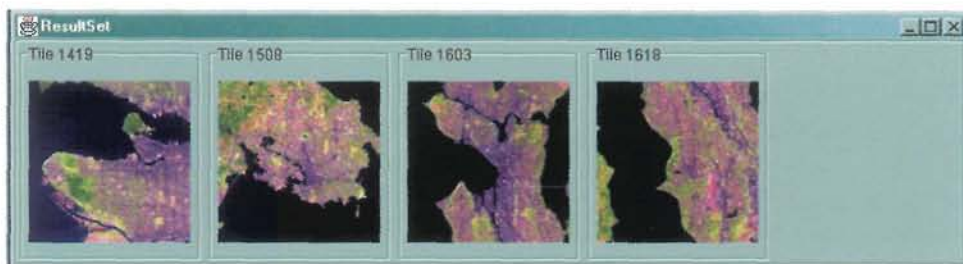


(b) Images classified as containing tree covered islands

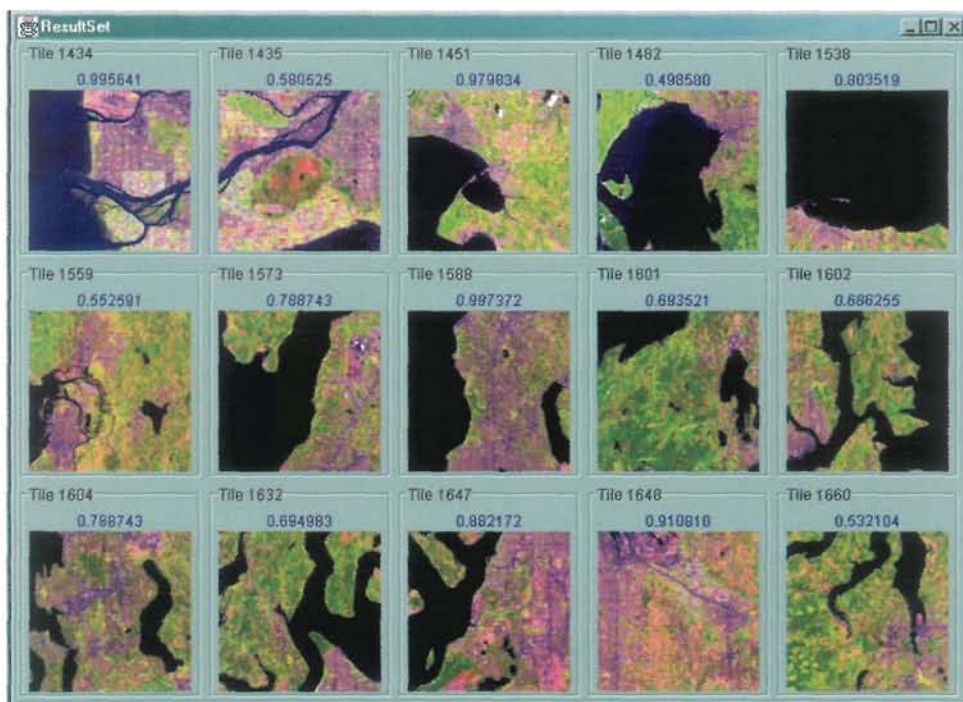
Fig. 15. Classification results for the “tree covered islands” class which is automatically modeled by the distinguishing relationships of green regions (lands covered with conifer and deciduous trees) surrounded by blue regions (water).

based on automatic selection of distinguishing (e.g. frequently occurring, rarely occurring) relations between regions.

The visual grammar overcomes the limitations of traditional region or scene level image analysis algorithms which assume that the regions or scenes consist of uniform pixel feature distributions. Furthermore, it can distinguish different interpretations of two scenes with similar regions when the regions have different spatial arrangements. We demonstrated our system with query scenarios that could not be expressed by traditional pixel, region or scene level approaches but where the visual grammar provided accurate classification and retrieval.



(a) Training images



(b) Images classified as containing residential areas with a coastline

Fig. 16. Classification results for the “residential areas with a coastline” class which is automatically modeled by the distinguishing relationships of regions containing a mixture of concrete, grass, trees and soil (residential areas) with their neighboring blue regions (water).

Future work includes using supervised methods to learn prototype models in terms of spectral, textural and ancillary GIS features; new methods for user assistance for updating of visual grammar models; automatic generation of metadata



for very large databases; and natural language search support (e.g. “Show me an image that contains a city surrounded by a forest that is close to a water source.”). Insightful Corporation’s INFACT product is a natural language question answering platform for mining unstructured data. A VISIMINE–INFACT interface will be an alternative to the query-by-example paradigm by allowing natural language-based searches on large remote sensing image archives. This will especially be useful for users who do not have query examples for particular scenes they are looking for, or when transfer of large image data is not feasible over slow connections.

## Acknowledgments

This work is supported by the NASA contract NAS5-01123 and NIH Phase II grant 2-R44-LM06520-02.

## References

1. S. Aksoy. *A Probabilistic Similarity Framework for Content-Based Image Retrieval*. PhD thesis, University of Washington, Seattle, WA, June 2001.
2. S. Aksoy and R. M. Haralick. Using texture in image similarity and retrieval. In M. Pietikainen, editor, *Texture Analysis in Machine Vision*, volume 40 of *Series in Machine Perception and Artificial Intelligence*, pages 129–149. World Scientific, 2000.
3. S. Aksoy and R. M. Haralick. Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern Recognition Letters*, 22(5):563–582, May 2001.
4. S. Aksoy, G. Marchisio, K. Koperski, and C. Tusk. Probabilistic retrieval with a visual grammar. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 2, pages 1041–1043, Toronto, Canada, June 2002.
5. S. Berretti, A. Del Bimbo, and E. Vicario. Modelling spatial relationships between colour clusters. *Pattern Analysis & Applications*, 4(2/3):83–92, 2001.
6. G. Celeux and G. Govaert. Gaussian parsimonious clustering models. *Pattern Recognition*, 28:781–793, 1995.
7. W. W. Chu, C.-C. Hsu, A. F. Cardenas, and R. K. Taira. Knowledge-based image retrieval with spatial and temporal constructs. *IEEE Transactions on Knowledge and Data Engineering*, 10(6):872–888, November/December 1998.
8. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., 2000.
9. C. Evans, R. Jones, I. Svalbe, and M. Berman. Segmenting multispectral Landsat TM images into field units. *IEEE Transactions on Geoscience and Remote Sensing*, 40(5):1054–1064, May 2002.
10. R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621, November 1973.
11. R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*. Addison-Wesley, 1992.
12. S. I. Hay, M. F. Myers, N. Maynard, and D. J. Rogers, editors. *Photogrammetric Engineering & Remote Sensing*, volume 68, February 2002.
13. G. G. Hazel. Object-level change detection in spectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(3):553–561, March 2001.
14. G. Koepfler, C. Lopez, and J. M. Morel. A multiscale algorithm for image segmentation by variational method. *SIAM Journal of Numerical Analysis*, 31:282–299, 1994.

15. K. Koperski, J. Han, and G. B. Marchisio. Mining spatial and image data through progressive refinement methods. *European Journal of GIS and Spatial Analysis*, 9(4):425–440, 1999.
16. K. Koperski, G. Marchisio, S. Aksoy, and C. Tusk. Applications of terrain and sensor data fusion in image mining. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 2, pages 1026–1028, Toronto, Canada, June 2002.
17. K. Koperski, G. Marchisio, S. Aksoy, and C. Tusk. VisiMine: Interactive mining in image databases. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 3, pages 1810–1812, Toronto, Canada, June 2002.
18. K. Koperski, G. Marchisio, C. Tusk, and S. Aksoy. Interactive models for semantic labeling of satellite images. In *Proceedings of SPIE Annual Meeting*, Seattle, WA, July 2002.
19. K. Koperski and G. B. Marchisio. Multi-level indexing and GIS enhanced learning for satellite images. In *Proceedings of ACM International Workshop on Multimedia Data Mining*, pages 8–13, Boston, MA, August 2000.
20. S. Kuehn, U. Benz, and J. Hurley. Efficient flood monitoring based on RADARSAT-1 images data and information fusion with object-oriented technology. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 5, pages 2862–2864, Toronto, Canada, June 2002.
21. K. I. Laws. Rapid texture classification. In *SPIE Image Processing for Missile Guidance*, volume 238, pages 376–380, 1980.
22. B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, August 1996.
23. G. B. Marchisio and J. Cornelison. Content-based search and clustering of remote sensing imagery. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 1, pages 290–292, Hamburg, Germany, June 1999.
24. G. B. Marchisio, K. Koperski, and M. Sannella. Querying remote sensing and GIS repositories with spatial association rules. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 7, pages 3054–3056, Honolulu, HI, July 2000.
25. G. B. Marchisio and A. Q. Li. Intelligent system technologies for remote sensing repositories. In C. H. Chen, editor, *Information Processing for Remote Sensing*, pages 541–562. World Scientific, 1999.
26. G. B. Marchisio, W.-H. Li, M. Sannella, and J. R. Goldschneider. GeoBrowse: An integrated environment for satellite image retrieval and mining. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 2, pages 669–673, Seattle, WA, July 1998.
27. G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. John Wiley & Sons, Inc., 1997.
28. P. J. Neal, L. G. Shapiro, and C. Rosse. The digital anatomist structural abstraction: A scheme for the spatial description of anatomical entities. In *Proceedings of American Medical Informatics Association Annual Symposium*, Lake Buena Vista, FL, November 1998.
29. E. G. M. Petrakis and C. Faloutsos. Similarity searching in medical image databases. *IEEE Transactions on Knowledge and Data Engineering*, 9(3):435–447, May/June 1997.
30. J. Pickrell. Aerial war against disease. *Science News*, 161:218–220, April 6 2002.
31. J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.
32. A. Rydberg and G. Borgefors. Integrated method for boundary delineation of agricul-

- tural fields in multispectral satellite images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(11):2514–2520, November 2001.
33. S. Santini. *Exploratory Image Databases: Content-Based Retrieval*. Academic Press, 2001.
  34. A. Sarkar, M. K. Biswas, B. Kartikeyan, V. Kumar, K. L. Majumder, and D. K. Pal. A MRF model-based segmentation approach to classification for multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 40(5):1102–1113, May 2002.
  35. M. Schroder, H. Rehrauer, K. Siedel, and M. Datcu. Interactive learning and probabilistic retrieval in remote sensing image archives. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5):2288–2298, September 2000.
  36. C.-R. Shyu, C. E. Brodley, A. C. Kak, and A. Kosaka. ASSERT: A physician-in-the-loop content-based retrieval system for hrct image databases. *Computer Vision and Image Understanding, Special Issue on Content-Based Access of Image and Video Libraries*, 75(1/2):111–132, July/August 1999.
  37. J. R. Smith and S.-F. Chang. VisualSEEk: A fully automated content-based image query system. In *Proceedings of ACM International Conference on Multimedia*, pages 87–98, Boston, MA, November 1996.
  38. L.-K. Soh and C. Tsatsoulis. Multisource data and knowledge fusion for intelligent SAR sea ice classification. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 1, pages 68–70, 1999.
  39. L. H. Tang, R. Hanka, H. H. S. Ip, and R. Lam. Extraction of semantic features of histological images for content-based retrieval of images. In *Proceedings of SPIE Medical Imaging*, volume 3662, pages 360–368, San Diego, CA, February 1999.
  40. L. H. Tang, R. Hanka, H. H. S. Ip, and R. Lam. Semantic query processing and annotation generation for content-based retrieval of histological images. In *Proceedings of SPIE Medical Imaging*, San Diego, CA, February 2000.
  41. J. C. Tilton, G. Marchisio, and M. Datcu. Image information mining utilizing hierarchical segmentation. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 2, pages 1029–1031, Toronto, Canada, June 2002.
  42. J. Xuan and T. Adali. Task-specific segmentation of remote sensing images. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, volume 1, pages 700–702, Lincoln, NE, May 1996.

## CHAPTER 4

### A SHAPE-BASED APPROACH TO CHANGE DETECTION AND INFORMATION MINING IN REMOTE SENSING

Jiang Li<sup>\*</sup>, Ram M. Narayanan<sup>\*</sup>, and Peter Z. Revesz<sup>†</sup>

*\*Department of Electrical Engineering  
University of Nebraska  
Lincoln, NE 68588-0511, USA  
E-mail: {jerry, ram}@doppler.unl.edu*

*†Department of Computer Science and Engineering  
University of Nebraska  
Lincoln, NE 68588-0115, USA  
E-mail: revesz@cse.unl.edu*

Shape analysis has not been considered in remote sensing as extensively as in other pattern recognition applications. However, shapes such as those of geometric patterns in agriculture and irregular boundaries of lakes can be extracted from the remotely sensed imagery even at relatively coarse spatial resolutions. This chapter presents a procedure for efficiently retrieving and representing the shape of objects in remotely sensed imagery using supervised classification, object recognition, and parametric contour tracing. Using the piecewise linear polygonal approximation technique, shape similarity can be compared by means of a computationally efficient metric. Our study was conducted on a time series of radiometric and geometric rectified Landsat Multispectral Scanner (MSS) images and Thematic Mapper (TM) images, covering the scenes containing lakes in the Nebraska Sand Hills region. The results show the effectiveness of our approach in detecting changes in lake shapes, which is potentially useful for specific applications such as the study of the lake change response to short or long term climatic variation and drought monitoring.

#### 1. Introduction

Shape retrieval and representation in remote sensing imagery have not received adequate attention as in other fields such as machine vision [1]. It is relatively easy to recognize the regular shapes, such as farm fields, which often appear as squares or rectangles, or highways or roads that are either straight or curved lines. However, irregular shapes such as lake boundaries tend to be scribbles and polygons without a set pattern. Recently, the point diffusion technique (PDT) [2], proposed to handle objects whose shape is not well defined and can be represented by set of sparse points, was applied to meteorological satellite images for fast and efficient shape similarity evaluation. Nevertheless, more attention needs to be given to the shape of the objects appearing in remotely sensed images. Shape analysis involves techniques such as image segmentation, object recognition, and contour extraction, etc. Other aspects that play an important role

in the algorithmic solutions are the approximation of shapes, the measurement of shape similarity, and the organization of shapes into search structures. An overview of the-state-of-art in shape analysis, classification, and similarity measure is given in [3] and [4]. Shape retrieval is generally carried out starting from an intermediate representation, typically involving the segmented image where the object shape has been located with special shape descriptors. Most popular image segmentation techniques are edge-based or region-based. Nevertheless, it is inherently difficult to extend differential methods to multi-band remotely sensed images by computing the gradient of the vector field [5] to obtain the edges of the regions. A hierarchical segmentation algorithm based on region growing and spectral clustering techniques was proposed in [6]. However, its practical implementation needs large-scale parallel processing to achieve reasonable processing times. Although hybrid approaches to integration of edge and region data in image segmentation exists [7], we adopted a supervised classification approach using the spectral information in the multi-band images to delineate the objects of concern.

Our study and analysis was conducted on a time series data set of radiometric and geometric rectified Landsat Multispectral Scanner (MSS) and Thematic Mapper (TM) images, covering scenes containing lakes in the Western Lakes Region of the Nebraska Sand Hills. The analysis of size, shape, and pattern of water-covered surfaces is potentially useful for flood and drought monitoring. MSS and TM images provide researchers with potentially unique chronological records that can be used to study surface water resources where no other hydrologic data exist. Buckwalter's [8] work stands as the original effort to monitor variations in the surface-water area of Sand Hills lakes. MSS data have been used to study the interrelationships between the timing of precipitation events and changes in the surface-water area of certain Sand Hills lakes [9], and to investigate the seasonal and interannual patterns of lake-size variability [10]. However, none of these studies addressed the issue of shape change of the lakes, which is important because lakes having little surface area variation may appear in much different shape and vice versa. Furthermore, the procedure for identifying and delineating surface water in these studies involves only simple level thresholding of the radiation signal in the near-infrared waveband, i.e., accepting scene pixels with low values while rejecting all others since the pixel value corresponding to water is typically low compared to other terrain features [11]. The challenge is to determine what specific level of the band reflectivity marks the "threshold" between terrestrial covers and standing surface water. Therefore, we applied a supervised classification technique using the support vector machine (SVM) in order to achieve more accurate delineated water



bodies. The SVM, a novel type of learning machine based on statistics learning theory introduced by Vapnik and Burges [12], [13], shows state-of-the-art performance in real-world applications including image classification. Recently, a study was conducted addressing the SVM-based classification scheme for land cover using polarimetric synthetic aperture radar (SAR) images [14]. In [15], it was concluded using hyperspectral AVIRIS images that the SVM outperforms the other traditional classification rules. A SVM-type classifier was developed for automatic classification of cloud data from GOES imagery in [16], and improved results were observed over other classification schemes such as Probabilistic Neural Networks.

We can mask out the water bodies in the classified images to obtain a series of binary images containing only the objects, i.e. lakes, of interest. To identify each individual object, we need to find the connected components in the binary image. There exist a number of algorithms for the connected components labeling operation. Tanimoto [17] assumes that the entire image can fit in memory and employs a simple recursive algorithm that works on one component at a time, but can move all over the image. Other algorithms were designed for larger images that may not fit in memory and work on only two rows of the image at a time. Rosenfeld and Pfaltz [18] developed the “classical” two-pass algorithm using a global equivalence table. Lumia, Shapiro, and Zuniga [19] developed another two-pass algorithm that uses a local equivalence table to avoid paging problems. In our work, we applied a connected components labeling algorithm with feature computation utility, adapted from the algorithm presented in [20]. The approaches to shape representation can be broadly divided into transform-based, region-based, and contour-based [4]. Transform-based methods such as Fourier or wavelets coefficients are frequently used as shape descriptors for classification. Representing shapes directly by binary images involves some drawbacks such as large storage space requirement. Therefore, we adopted the contour-based approach, which can be further divided into nonparametric contours in which the shape outline is simply represented as a set of points, without any special order among them, and parametric contours in which the shape outline is represented as a parametric curve implying a sequential order along it. Parametric contours can be simplified and compressed by curve approximation technique, e.g., using a set of geometric primitives such as straight-line segments or splines to fit the shape outline. The contours need to be converted into polygons before a shape similarity metric can be applied to quantify the shape change. Meanwhile, by polygonal approximation, the spur pixels introduced by classification errors can be eliminated to a certain degree and the total data required to be stored can be significantly reduced. The

polygonal approximation problem has been studied by many researchers [21]-[23]. The ideal procedure is to represent lines and boundaries by means of polygons with minimum number of vertices under a certain fit criterion. It remains an open problem to find a piecewise linear interpolation algorithm that both returns the fewest number of pieces and runs in  $O(n)$  time. In this research, we developed a polygonal approximation algorithm running in  $O(n)$  time, adapted from the piecewise linear interpolation algorithm for time series data presented in [24] and [25]. This algorithm produces a small but not minimum number of vertices that lie on the given contour described by an array of points. In order to find out the similarity between the shapes of a lake at different dates, we resorted to an efficiently computable shape similarity metric proposed in [26].

In this chapter, we introduce our approach to efficiently retrieve and represent interesting shapes, e.g. lakes, in remotely sensed imagery. After discussing the supervised classification using support vector machines and comparing the classification accuracy with other classifiers in Section 2, a description of connected components labeling and parametric contour tracing is presented in Section 3. The piecewise linear polygonal approximation algorithm is developed in Section 4. Section 5 presents the computable similarity metric for measuring the similarity between shapes, and Section 6 presents the study results with further discussion. Finally, Section 7 contains conclusions and proposals for future work.

## 2. Supervised Image Classification

The imagery consists of 36 4-band Landsat MSS images ( $256 \times 256$  pixels), covering scenes containing the lakes of the Nebraska Sand Hills region encompassing the period from 1981 to 1987, and 10 7-band Landsat TM images ( $768 \times 768$  pixels) covering the same area from 1992 to 1997. These images have been pre-calibrated and registered to UTM-13 map. As simple level thresholding technique is challenged by the requirement of accurately separating the water bodies. Therefore, we adopted a supervised classification approach using Support Vector Machine. Classification of remote sensing images using SVM has been reported to be computationally simple and can result in accuracy better than other more computationally intensive rules [14] – [16].

The training and test data were sampled by a region growing algorithm. The algorithm simultaneously sets the spatial constraints including the maximum area  $A$  of the region and the maximum distance  $D$  from the seed pixel, and spectral Euclidean distance  $S$ , which refers to those pixels whose spectral reflectance is within a certain distance from the mean of the region pixels. The basic procedure

is outlined in Table 1, where  $(r_s, c_s)$  is the coordinate of the seed pixel,  $I(r_n, c_n)$  is the gray level of a pixel from region  $R$  with coordinates  $(r_n, c_n)$  and  $\mu_R$  is the mean gray level of all pixels in  $R$ .

Table 1. Region growing algorithm for training data sampling

```

repeat until  $A_R > A$ 
for each pixel  $p$  at the border of  $R$  do
  for all neighbors  $p_n$  at  $(r_n, c_n)$  of  $p$  do
    if  $\|(r_n, c_n) - (r_s, c_s)\| \leq D$  and  $\|I(r_n, c_n) - \mu_R\| \leq S$  then
      add  $p_n$  to  $R$  and update  $\mu_R$ 
    end if
  end for
end for

```

According to a general rule given in [27], more than  $10n$  pixels of training data should be sampled for each class where  $n$  is the number of bands. Hence, we sampled 100 pixels for each class in a MSS image and 200 pixels in a TM image. Five data sets for each class were sampled at different sites and we used 20% of the sampled data for training and the remaining 80% for testing.

### 2.1. Theoretical Background of SVM

A comprehensive introduction to SVM is given in [28] and we briefly review the basics as follows. Given a set of examples consisting of pairs of class labels and  $n$ -dimensional feature vectors as  $(y_i, \mathbf{x}_i), i = 1, \dots, l, y_i \in \{1, -1\}, \mathbf{x}_i \in \mathbf{R}^n$ , we want to find a decision function  $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b, \mathbf{x} \in \mathbf{R}^n, b \in \mathbf{R}$  satisfying  $y_i \cdot \text{sgn}(f(\mathbf{x}_i)) > 0$  in the linearly separable case. The SVM approach places the hyperplane  $\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$  so that the margin, which is defined as the distance of the closest vectors in both classes to the hyperplane, is maximized. It can be shown that the geometric margin is computed as  $1/\|\mathbf{w}\|_2$  and the corresponding hyperplane is obtained by the optimization problem:

$$\min_{\mathbf{w}, b} \langle \mathbf{w}, \mathbf{w} \rangle, y_i (\langle \mathbf{w}_i, \mathbf{x}_i \rangle + b) \geq 1, i = 1, \dots, l. \quad (1)$$

This optimization problem can be translated into the following form by introducing the Lagrange multipliers  $\alpha_i \geq 0$ , subject to  $\sum_{i=1}^l \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, \dots, l$ .

$$\max W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l y_i y_j \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle, \quad (2)$$

Only a small number of multipliers  $\alpha_i$  have nonzero values and they are associated with the exemplars, or the so-called support vectors, which form the boundaries of the classes. This is significant because it is usually true that a small subset of all the training data are actually involved in defining the hyperplane,

i.e., those examples that are closest to the hyperplane. The maximal margin classifier can be generalized to nonlinearly separable data via two approaches. One is to introduce a soft margin parameter  $C$  to relax the constraint that all the training vectors of a certain class lie on the same side of the optimal hyperplane. This approach is effective in case of noisy data. The other is to transform input vectors into a higher dimensional feature space by a map function  $\phi$  and then do a linear separation there. The expensive computation of inner products  $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$  can be reduced significantly by using a suitable kernel function  $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ . That is, we do not need to have an explicit representation of  $\phi$ , but only  $K$ .






## 2.2. Structure of the Classifier

We implemented the Support Vector Machine classifier using the SVM LIB library [29], which is based on both Sequential Minimal Optimization (SMO) algorithm by Platt [28] and SVM<sup>Light</sup> by Joachims [30]. The feature vector is composed of the intensity in each of the four bands  $[I_1, I_2, I_3, I_4]$  for MSS data and in each of the six bands  $[I_1, I_2, I_3, I_4, I_5, I_7]$  for TM data, in which the elements are standardized to have the same range, i.e.,  $[0, 1]$ . The one against one method, which constructs  $k(k-1)/2$  hyperplanes that discriminate each pair of classes, was chosen to construct the multi-class classifier. The library also provides a set of optional kernel functions that satisfy the necessary and sufficient conditions for the expansion of the kernel function given by Mercer's theorem [31]. Radial basis function (RBF) defined in (3) below is used as the kernel to build each of the binary classifier.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (3)$$

The soft margin parameter  $C$  was decided by the cross validation function offered by the library. We set  $C = 100$  after trying a series of values ranging from 1 to 500. We choose  $\gamma = 0.5$  as in [32], wherein it is reported that the SVM is not sensitive to different choices of parameter  $\gamma$  for RBF. According to USGS Land Use / Land Cover Classification scheme and the land cover map produced by Nebraska GAP project, five land cover classes are defined as shown in Table 2. Fig. 1 (a) and Fig. 1 (b) show the original TM image and the classified image in pseudo color respectively.

Table 2 Land cover categories

Land Cover Type	Color
Open Water	
Emergent Wetlands	
Bare Rock/Sand/Clay	
Grasslands/Herbaceous	
Pasture/Hay	

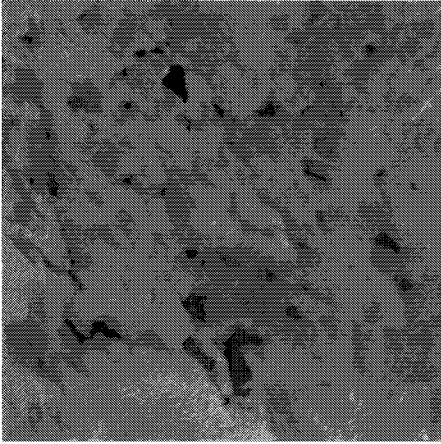


Fig. 1 (a). Original Landsat TM image.

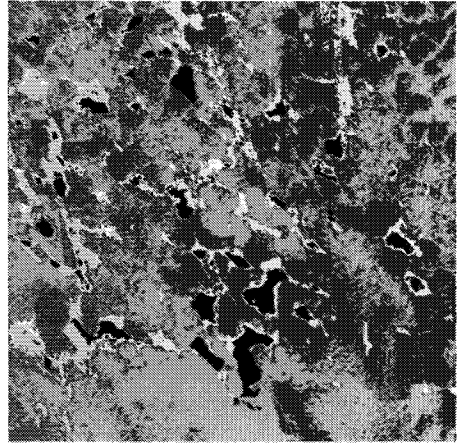


Fig. 1 (b). Classified Landsat TM image.

### 2.3. Accuracy Assessment

The classification was repeated five times using a different 20% subset of the sampled data for training and using the remaining 80% for test. Table 3 gives an average performance for different classes for a typical MSS image and a typical TM image. Note both the producers accuracy and users accuracy are very high and water was classified almost completely correctly. In general, results obtained from the TM image are better than those from the MSS image. Table 4 compares the overall classification accuracy of the Support Vector Machine classifier, the Maximum Likelihood (ML) classifier, and the Minimum Distance to Means (MDM) classifier. It shows that the SVM classifier outperforms both the ML classifier and the MDM classifier.

Table 3. Classification accuracy for different classes

Class Name	MSS Image		TM Image	
	Producer's Accuracy	User's Accuracy	Producer's Accuracy	User's Accuracy
Open Water	98.4%	98.7%	99.8%	99.7%
Emergent Wetlands	91.7%	92.6%	93.3%	93.6%
Bare Rock/Sand/Clay	89.5%	87.8%	90.4%	88.1%
Grasslands/Herbaceous	88.3%	86.4%	89.0%	86.8%
Pasture/Hay	92.1%	94.5%	94.9%	96.4%

Table 4. Overall performances for different classifiers

Classifier	Overall Classification Accuracy	
	MSS Images	TM Images
Support Vector Machine	91.5%	93.6%
Maximum Likelihood	87.2%	89.8%
Minimum Distance to Means	80.6%	82.3%

### 3. Object Recognition and Parametric Contour Tracing

The water bodies were masked out from the classified image to obtain a binary image containing all the objects (lakes of interest) as shown in Fig. 2 (a). Our task now is to recognize each of the objects using a connected components labeling algorithm that searches the entire image to find all the different sets of pixels that are connected to each other. The labeling algorithm used in this research was modified from the one introduced in [20], which is briefly described here. The definition of the connected component is as follows [20]. Suppose two pixels occur in a binary image where  $I(r,c) = I(r',c') = v$ ,  $v \in \{0, 1\}$ . The pixel  $(r,c)$  is connected to the pixel  $(r',c')$  with respect to value  $v$  if there is a sequence of pixels  $S : (r,c), (r_0,c_0), \dots, (r_n,c_n), (r',c')$  in which  $I(r_i,c_i) = v$ ,  $i = 0, \dots, n$ , and  $(r_i,c_i)$  is the neighbor of  $(r_{i-1},c_{i-1})$ . The sequence of pixels  $S$  forms a connected path from  $(r,c)$  to  $(r',c')$ . A connected component of value  $v$  is a set of pixels  $C$ , each having value  $v$ , and such that every pair of pixels in the set is connected with respect to  $v$ . The algorithm scans through a binary image and locates connected regions defined by "1" pixels against a background of "0" pixels. The assumption is that each region of "1"s represents the image of a single object. When this algorithm finishes labeling all the "1" pixels (object) reachable from the first found "1", it then creates another label and searches for the next unlabeled "1" pixel and propagates anew. The labeling process is finished when all the "1" pixels in the image are exhausted.

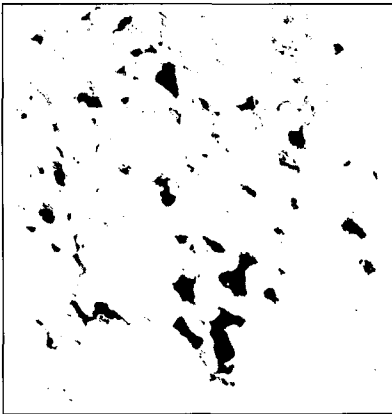


Fig. 2 (a). Binary image containing water.

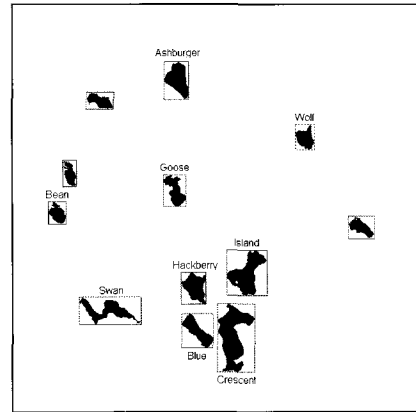


Fig. 2 (b). Lakes of interest with Feret box.

The area of each object is calculated by counting the number of pixels of each connected components during the labeling. In this study, we only consider the lakes with average size larger than a threshold, for instance, 200 for MSS images and 1000 for TM images, so that the retrieval and representation of shape are nontrivial. Therefore, tiny objects can be removed as nonlake objects or just

classification errors. The objects with the corresponding Feret box, defined as the smallest rectangle that orients with respect to the coordinate axis and encloses the object, are computed during the labeling as shown in Fig 2 (b), wherein the name of each studied lake is indicated. The histogram of the average area for each identified object, as shown in Fig. 3, was plotted to assist the threshold selection.

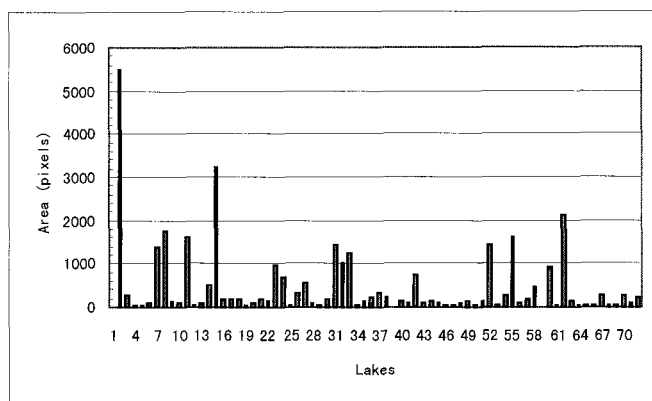


Fig. 3. Histogram of the average area of each lake.

The procedure of obtaining a parameterized representation of the contour is referred to as contour tracing, or alternately contour following and contour extraction in some literature. There are many algorithms for extracting parametric contours in a binary image, and some examples include the use of run-length codes [33] and chain-codes [34]. However, the assumption of these algorithms is that only one object exists in a given image. In our case, the binary image contains several objects. To avoid manually discerning each of the lakes in the connected components, we developed an algorithm that automatically goes through all the Feret boxes associated with the lakes to extract their contours and features such as centroids and perimeters. Our algorithm is based on an efficient contour following algorithm proposed in [4], which can trace out the external contour of a single connected object in a binary image and join them into a list of points. The initial point belonging to each of the objects can be found using raster scanning, i.e., search line by line, until the first “0” pixel having a labeled pixel as a neighbor at its right is found. As shown in Fig. 4., the neighbors at the positions 0, 1, 2, 3 have already been visited during the scan. Fig. 5. illustrates the possible positions of four candidates, i.e., 4, 5, 6, 7. After identifying the start pixel of the object, the algorithm circumnavigates the connected object until the starting pixel is revisited, which indicates the completion of the tracing of the current object. The pseudo-code of this algorithm is presented in [4], which will not be further discussed here.

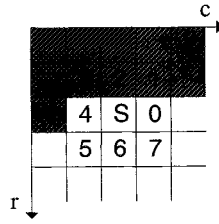


Fig. 4. Verified positions during scanning search.

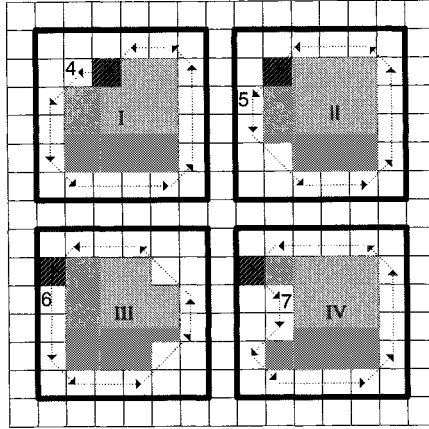


Fig. 5. Automatic contour tracing on connected components.

Since the Feret boxes might overlap with each other, the tracing algorithm regards the pixels with a specific component label as the object and other pixels as background within a given Ferex box. The location of each of the lakes of interest can be automatically recognized by referring to the UTM coordinates given in [10]. Specifically, we converted the UTM coordinates of each lake to the relative coordinates with a binary image by mapping the UTM coordinate of the pixel at the upper-left corner to origin. Suppose the retrieved contour is represented as a complex-valued signal  $c[n] = x[n] + iy[n]$ , with  $n = 0, \dots, N-1$ , the centroid of the lake can be obtained as the average value of all points of  $c[n]$ , i.e.,

$$C = \frac{\sum_{n=0}^{N-1} c[n]}{N} \quad (4)$$

We then calculate the Euclidean distance between the centroid and each of the reference coordinates and assign the name of the lake with the one having the minimum distance. The perimeter of each lake can be computed according to 8-neighborhood connectivity, i.e., the length between consecutive pixels is  $\sqrt{2}$  if the pixels are diagonal neighbors, and 1 otherwise. The perimeter is computed as



$$P = \sum_{n=0}^{N-1} \|c[n+1] - c[n]\|, \quad (5)$$

where  $c[N] = c[0]$  and  $\|\cdot\|$  denotes the complex modulus.

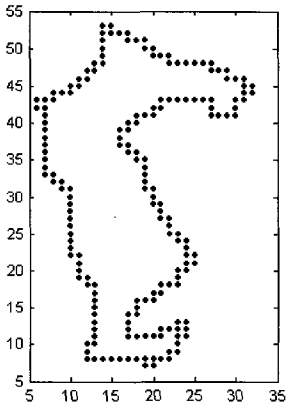
The complete algorithm is summarized in Table 5. The retrieved contours of Crescent Lake on different dates are shown in Fig. 6 (a) to Fig. 6 (c), in which we can observe the noticeable variation of the shape of the lake over the two-year period.

Table 5. Parametric contour Tracing and lake recognition

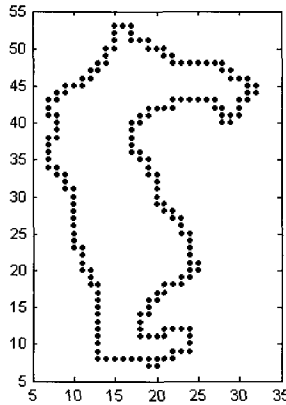
```

for each Feret box  $F_i$  do
  Trace and save the object's contour
  Compute centroid  $C_i$  and perimeter  $P_i$  of the object
  for each lake with reference coordinate  $RC_j$  do
     $d_{ij} = \text{Euclidean\_Distance}(C_i, RC_j)$ 
    Assign object the name of the lake with  $\min(d_{ij})$ 
  end for
end for

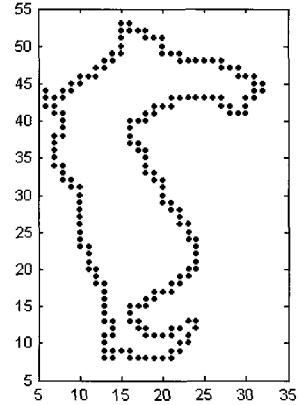
```



(a) June 1983.



(b) June 1984.



(c) June 1985.

Fig. 6. Retrieved contours of Crescent Lake.

#### 4. Piecewise Linear Polygonal Approximation

The contour needs to be converted into a polygonal shape before a shape similarity metric can be applied so as to a quantitative measurement of the lake boundary change. Using polygonal approximation, noise introduced by spur pixels due to the classification errors can be reduced and so does the total data that needs to be stored. The ideal procedure is to represent the contour by means of polygons with a minimum number of vertices and satisfying a given fit

criterion. There has been much work carried out on the polygonal approximation problem using the piecewise linear functions [21]-[23]. However, it remains an open problem to find a piecewise linear interpolation algorithm that both returns the fewest number of pieces and runs in  $O(n)$  time. In this research, we developed a piecewise linear polygonal approximation (PLPA) algorithm running in  $O(n)$  time, adapted from the piecewise linear interpolation algorithm for time series data presented in [24], [25]. This approximation algorithm produces a small but a non-minimum number of vertices that lie on the given contour presented by an array of points.

Given a sequence of points  $S : (x_0, y_0), (x_1, y_1), \dots, (x_{N-1}, y_{N-1})$  and an error threshold  $\Psi$ , PLPA finds a piecewise linear function  $f$ , composed of a set of segments represented by linear functions  $f_j$ ,  $j=1, \dots, K$  where  $K$  is the total number of pieces, whose domains are disjoint. Assuming that the start and end points of a piece are  $[x_m, y_m]$  and  $[x_n, y_n]$  respectively as shown in Fig. 7 (a), each linear function  $f_j$  is defined as

$$f_j(x) = \frac{y_n - y_m}{x_n - x_m} (x - x_m) + y_m. \quad (6)$$

The lower line and upper line according to the threshold  $\Psi$  are denoted as:

$$L_j(x) = \frac{(y_n - \Psi) - y_m}{x_n - x_m} (x - x_m) + y_m \quad (7)$$

$$U_j(x) = \frac{(y_n + \Psi) - y_m}{x_n - x_m} (x - x_m) + y_m \quad (8)$$

PLPA tries to extend each piece until it cannot because the upper line  $U_j$  becomes lower than the lower line  $L_j$ . Fig. 7 (b) illustrates the case  $x_n < x_m$ , in which  $L_j$  now starts from  $(x_m, y_m)$  and passes through  $(x_n, y_n + \Psi)$  while  $U_j$  starts from the same point and passes through  $(x_n, y_n - \Psi)$ .

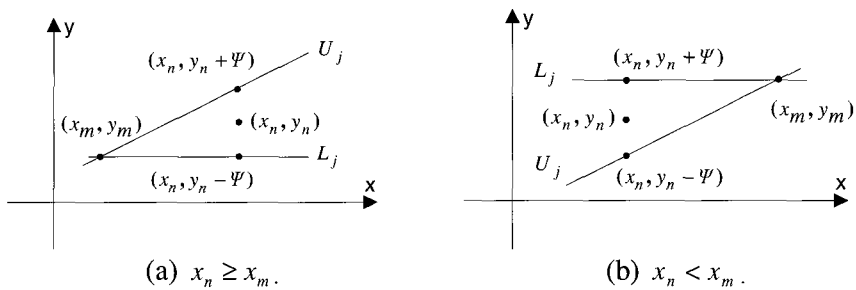


Fig. 7. Start and end points of a segment.

Table 6 shows the algorithm in which the input is the point sequence  $S$ , i.e., the coordinates of the contour, and a maximum error threshold  $\Psi$ , the output is a piecewise linear polygonal approximation function. The variable  $Start$  is the start point with  $S_L$  and  $S_U$  representing the slope of a piece. The proof of the correctness of this algorithm is similar to those given in [25] and not presented here.

Table 6. Piecewise linear polygonal approximation algorithm

```

Start = ( $x_1, y_1$ ),  $S_L = -\infty$ ,  $S_U = +\infty$ 
for  $i=1$  to  $n-1$  do
     $S'_L = \max(S_L, \text{slope}(Start, (x_{i+1}, y_{i+1} - \Psi)))$ 
     $S'_U = \min(S_U, \text{slope}(Start, (x_{i+1}, y_{i+1} + \Psi)))$ 
    if  $x_{i+1} \leq x_i$  then
        Exchange( $S'_L, S'_U$ )
    end if
    if  $S'_L \leq S'_U$  then
         $S_L = S'_L$ ,  $S_U = S'_U$ 
    else
         $f_i(x) = \frac{S_L + S_U}{2}(x - Start.x) + Start.y$ 
         $Start = (x_i, f_i(x_i))$ 
         $S_L = \text{slope}(Start, (x_{i+1}, y_{i+1} - \Psi))$ 
         $S_U = \text{slope}(Start, (x_{i+1}, y_{i+1} + \Psi))$ 
        if  $x_{i+1} \leq x_i$  then
            Exchange( $S_L, S_U$ )
        end if
    end if
end for

```

The approximated polygons of the contour of Crescent Lake in June 1985 under different thresholds are shown in Fig. 8 (a) to Fig. 8 (c). It can be seen that the smaller  $\Psi$  is, the more segments the piecewise linear function (i.e., the polygon) will contain, which preserves more details of the shape.

The compression ratio (CR), defined in (9) below, for different thresholds is shown in Fig. 9 (a), which shows that the larger the threshold value, more the number of points along the contour can be eliminated.

$$CR = 1 - \frac{\# \text{ of polygon vertices}}{\# \text{ of contour points}} \quad (9)$$

Since the threshold measures the maximal distance between the original points and the approximated line, theoretically, suppose the length of the perimeter is  $P$ , the upper-bound of the approximation error in the area will be  $e_A \leq P \cdot \Psi$ . Hence lakes with high perimeter to area ratio ( $PAR$ ) defined as  $P/A$  will have higher errors in general than those lakes that have low  $PAR$ . We test this on different representative lakes and Fig. 9 (b) shows the approximation precision ( $AP$ ), defined in (10) where  $A$  represents the original area and  $\tilde{A}$  is the approximated value, for different compression ratios. Note that very high precision is conserved even at 55% compression ratio. Therefore, we chose 0.25 as the threshold.

$$AP = 1 - \left| \frac{A - \tilde{A}}{A} \right| \tag{10}$$

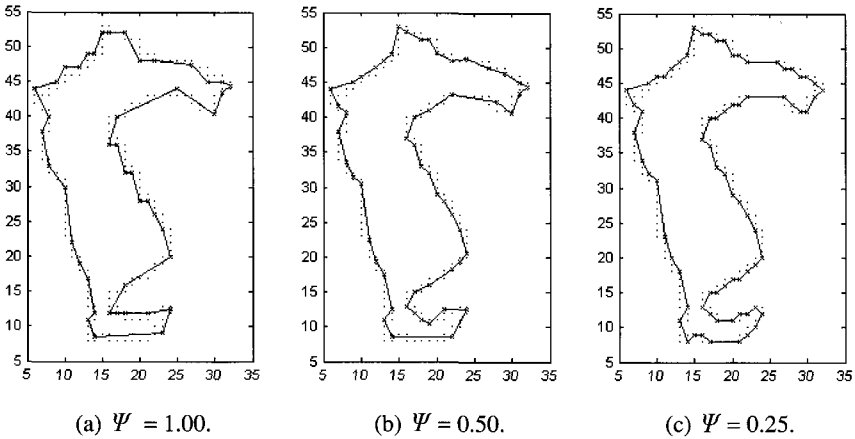


Fig. 8. Approximated polygons of the contour of Crescent Lake in June 1985 under different thresholds.

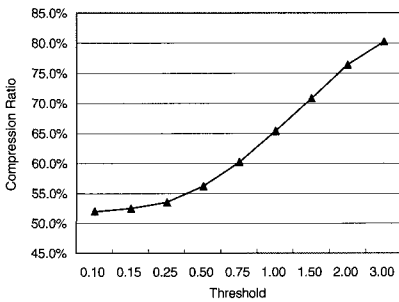


Fig. 9 (a). Compression ratio vs. thresholds.

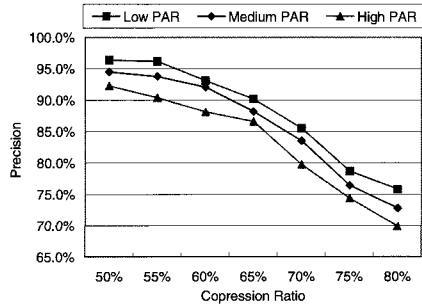


Fig. 9 (b). Precision vs. compression ratios.

## 5. Shape Similarity Measure

Assuming that the noise generated by the classification error is roughly uniformly distributed over the sides of the approximated polygon, we can apply a shape similarity metric discussed below to give a quantitative measure of the change of a shape, i.e., to evaluate the similarity between the shapes of a specific lake on different dates. A similarity measure is a function defined on pairs of shapes indicating the degree of resemblance of the shapes and it is desirable that such a similarity measure be a metric. Suppose  $S$  is any set of shape objects, a metric on  $S$  is a function  $d(.,.): S \times S \rightarrow \mathfrak{R}$  satisfying the following properties for all  $X, Y, Z \in S$  [3]:

$$d(X, X) = 0 \quad (11)$$

$$d(X, Y) = 0 \Rightarrow X = Y \quad (12)$$

$$d(X, Y) = d(Y, X) \quad (13)$$

$$d(X, Y) + d(Y, Z) \geq d(X, Z) \quad (14)$$

In this research, we used an efficiently computable metric proposed in [26]. This metric is based on the  $L_2$  distance between the turning functions of the two polygons that can be both convex and nonconvex. It has several important properties. First, it is invariant under translation, rotation, and change-of-scale. Second, it can be computed in time  $O(mn \log mn)$  where  $m$  is the number of vertices in one polygon and  $n$  is the number of vertices in the other. Third, it is insensitive to small perturbations and matches human intuitive notions of shape resemblance. The detailed representation of polygons and corresponding distance function were presented in [26] and briefly described as follows.

### 5.1. Polygon Representation by Turning Function

A polygon  $A$  can be represented by a list of vertices around its boundary with corresponding coordinates. An alternative method is to use the turning function  $\theta_A(s)$ , which measures the cumulative angle of the counterclockwise tangent as a function of the arc length  $s$ , measured from some reference point  $O$  on  $A$ 's boundary as shown in Fig. 10. It increases with left-hand turns and decreases with right-hand turns. The polygon is rescaled so that the total perimeter length is 1, i.e.,  $s \in [0,1]$ , and  $\theta_A(s)$  is a function from  $[0,1]$  to  $\mathfrak{R}$ . Since  $\theta_A(s)$  may become arbitrarily large over  $[0,1]$  for a nonconvex polygon, we should have  $\theta_A(1) = \theta_A(0) + 2\pi$  in order to represent a closed polygon [26].

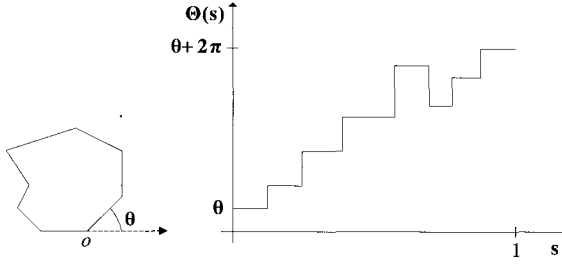


Fig 10. Polygon representation using turning function.

## 5.2. Distance Measure

Given two polygons  $A$  and  $B$  and their associated turning functions  $\Theta_A(s)$  and  $\Theta_B(s)$ , we can measure the degree to which they are similar by taking the distance between  $\Theta_A(s)$  and  $\Theta_B(s)$  in terms of the metric on function spaces. The  $L_p$  distance between  $A$  and  $B$  is defined as [26]

$$\delta_p(A, B) = \|\Theta_A - \Theta_B\|_p = \left( \int_0^1 |\Theta_A(s) - \Theta_B(s)|^p ds \right)^{\frac{1}{p}}, \quad (15)$$

where  $\|\cdot\|_p$  denotes the  $L_p$  norm.

Suppose the turning function is  $\Theta_A(s+t)$  if the reference point  $O$  along  $A$ 's boundary is shifted by  $t$ , and it is  $\Theta_A(s)+\theta$  if  $A$  is rotated by  $\theta$ , then to minimize the distance over all such effects, the similarity measure in case of the planar polygons, i.e.,  $P = 2$ , is given as [26]

$$d_2(A, B) = \left( \min_{\theta \in \mathbb{R}} \min_{t \in [0,1]} \int_0^1 |\Theta_A(s+t) - \Theta_B(s) + \theta|^2 ds \right)^{\frac{1}{2}}. \quad (16)$$

## 6. Study Results

Lake change is characterized by its variations of area, perimeter, and shape. Reference [9] has reported that generally the lakes in this study area experience surface-area maxima in June with minima occurring in October, which is verified in our research. However, we adopted relative values instead of absolute values. Results obtained from MSS data set are shown in Fig. 11 to Fig 13. Fig. 11 (a) shows the variability of the area occurred in consecutive months for the studied lakes in 1983. Note that the area of Goose Lake had a considerable increase during May to June and significant decrease during August to September. The corresponding shape similarity measures are shown in Fig. 11 (b), in which the smaller the measurement is, the more similar the compared two shapes are. It can

be observed that the shape variation is not consistent with area fluctuation. Island, Hackberry, and Goose's shapes show more variation than those of other three lakes and the shape of Goose Lake presents the largest variability.

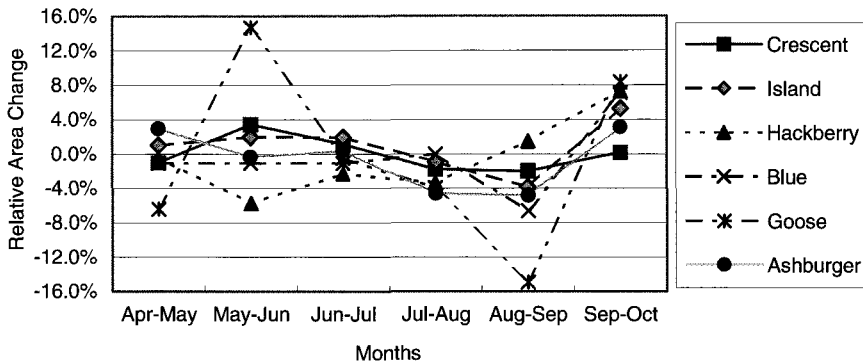


Fig. 11 (a). Seasonal variability of the area in 1983.

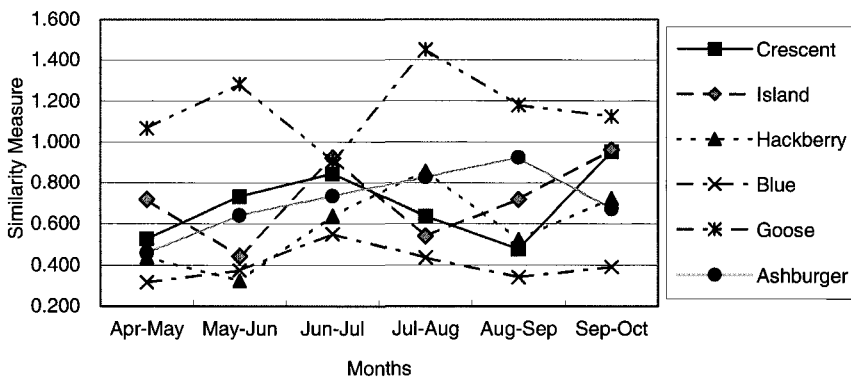


Fig. 11 (b). Shape similarity measure in 1983.

The overall fluctuation is examined by the standard deviation of the relative area change occurred in each year as shown in Fig. 12. It is observed that the fluctuation in area is significant during May to June and August to October. We calculated coefficient of variation  $CV = \sigma / \mu$  [10] for the area and the perimeter for all the lakes in each year, where  $\mu$  is the mean and  $\sigma$  is the standard deviation. The average values are shown in Fig. 13. Note that Island Lake and Goose Lake have  $CV \geq 10\%$  for both area and perimeter. An interesting discovery is that the  $CV$  of perimeter is always less than that of area except for Goose Lake, which might be caused by its more irregular shape.

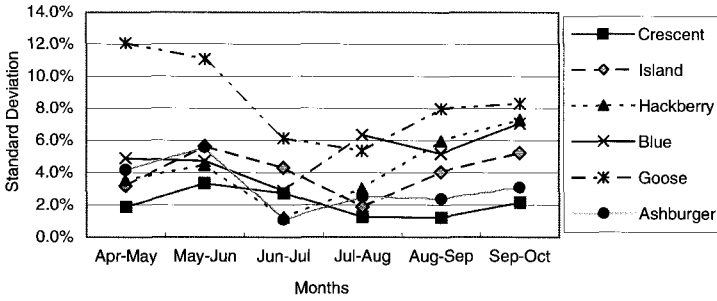


Fig. 12. Standard deviation of the relative area change.

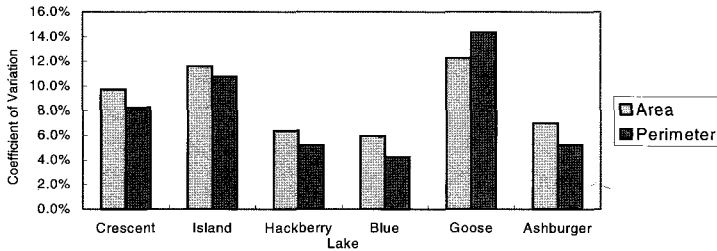


Fig. 13. Average coefficient of variation of lake area and perimeter.

It has been found that in our study area, a good correlation exists between lake-surface area and precipitation occurring over the previous 45 days, and this pattern tends to be repeatable and steady [9]. Now we are interested in the relationship between the lake changes and the drought events. Specifically, the area variations for each lake were correlated with the short-term (1-month) drought indices. Although there exist many kinds of drought indices, we employed the modified Palmer Drought Severity Index (PDSI), which has been used in the National Drought Atlas, as a good representation of existing conditions for real-time, operational use [35]. Fig. 14 illustrates the variation of the PDSI in different years. Drought is defined as beginning in the month when the PDSI equals or falls below  $-1$  after having been above  $-1$ , and the drought duration is defined as the interval of time for which the PDSI remains equal or below  $-1$  [35]. Fig. 15 shows the correlation coefficient between the areas and the indices for each of the lake. Strong correlations are observed although the coefficients are not always positive. It shows positive correlation when the indices stay below  $-1$ , i.e., within the drought period as in 1981 and 1987. Negative correlations are observed at end of the drought duration, i.e., the indices go above  $-1$  as in 1983 and 1984. Note a very strong correlation occurred in 1985 when the indices are above  $-1$ , which indicates a nondrought period.



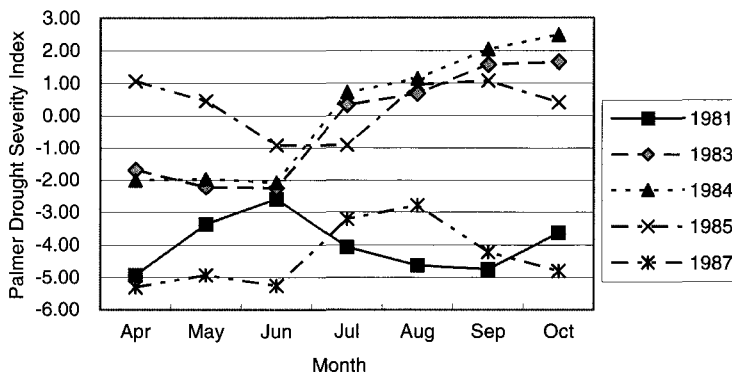


Fig. 14. Variation of Palmer Drought Severity Index.

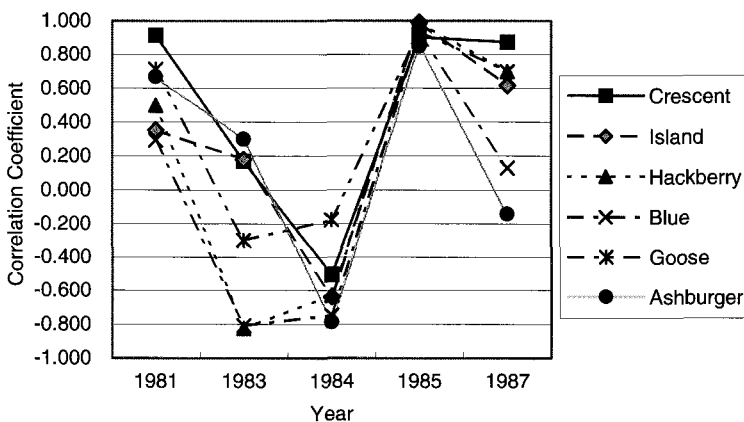


Fig. 15. Lake area / Palmer Drought Severity Index correlations.

Since TM data has a higher spatial resolution, we studied three more lakes, i.e., Wolf, Bean, and Swan. However, due to the fact that the total number of the TM images available to us was small, results obtained from TM data set only illustrate changes occurring during summer (June vs. August) and autumn (September vs. October) as shown in Fig. 16 to Fig. 18. It can be seen from Fig. 16 and Fig. 17 that Crescent, Island, Hackberry, and Blue always have no significant area change except in 1993, in which all the lakes show dramatic area reduction, except Wolf who showed an increase. All the lakes show the nearly same area fluctuation pattern in fall except Ashburger who shows a considerable gain and Bean who shows a remarkable loss. With regard to perimeter, Goose and Bean show substantial variability while other lakes have similar pattern in fall. Note that Swan demonstrates more fluctuations of both area and perimeter than other lakes in summer but not in autumn. It also can be observed that significant area change does not always imply large perimeter change.

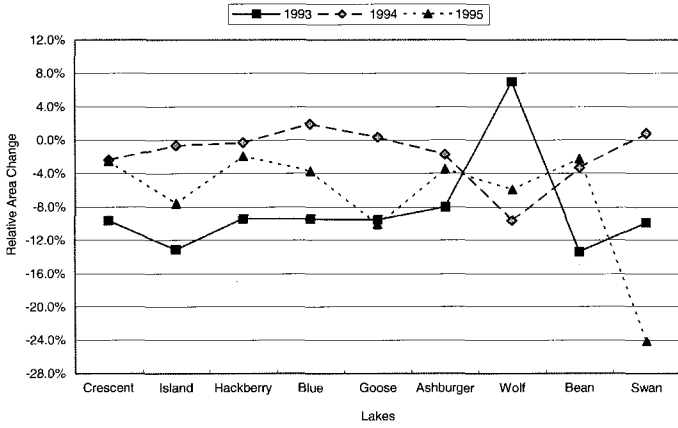


Fig. 16 (a). Surface-area fluctuations of the lakes in Summer.

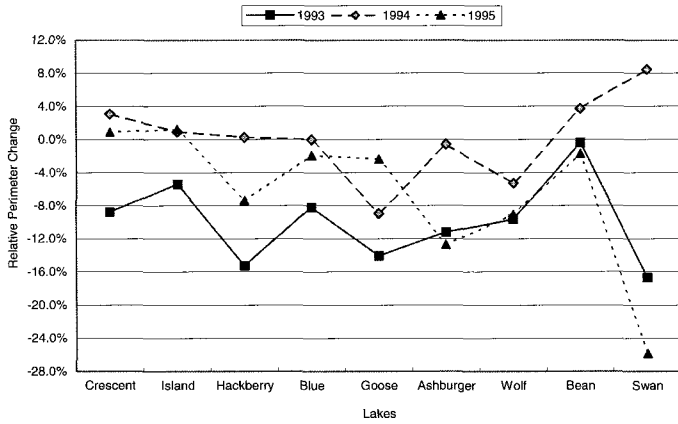


Fig. 16 (b). Perimeter variability of the lakes in Summer.

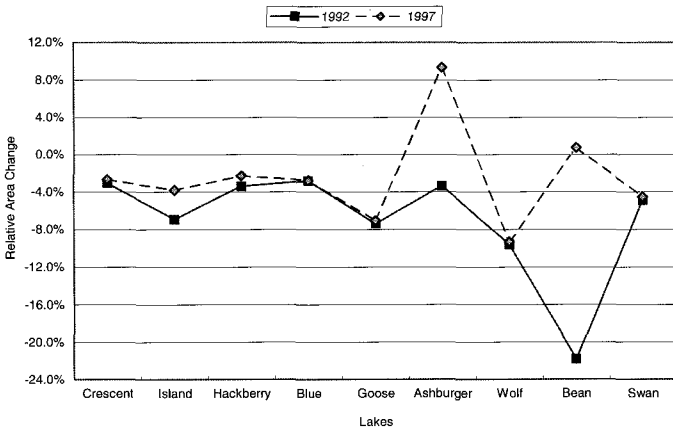


Fig. 17 (a). Surface-area fluctuations of the lakes in Autumn.

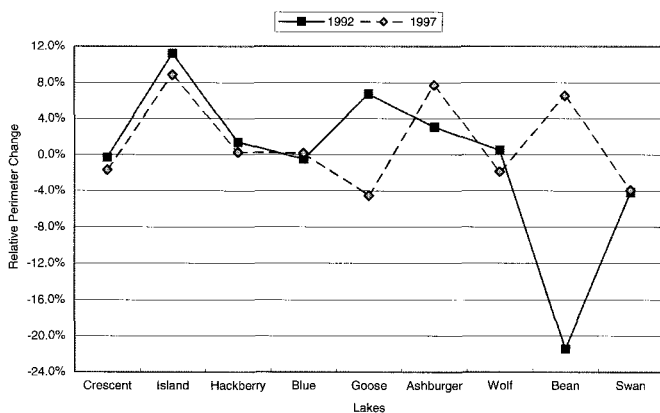


Fig. 17 (b). Perimeter variability of the lakes in Autumn.

It is observed from Fig. 18 that the pattern of the shape similarity in summer is more obvious than that in autumn. Blue and Hackberry always have their shape preserved as the similarity measure is around 0.5, except that Hackberry attained 1.4 in summer 1995. Ashburger and Wolf show shape similarity variability from about 1.0 in summer to about 0.7 in autumn. Note the variability of the shape of Island in summer is larger than that in autumn and Goose's shape changes dramatically both in summer and autumn.

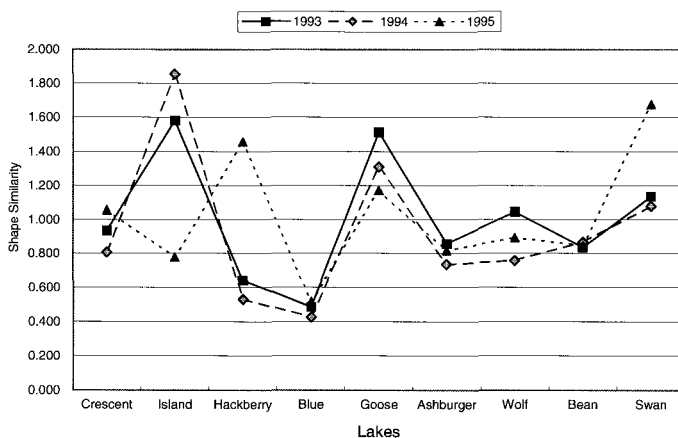


Fig. 18 (a). Shape similarity of the lakes in Summer.

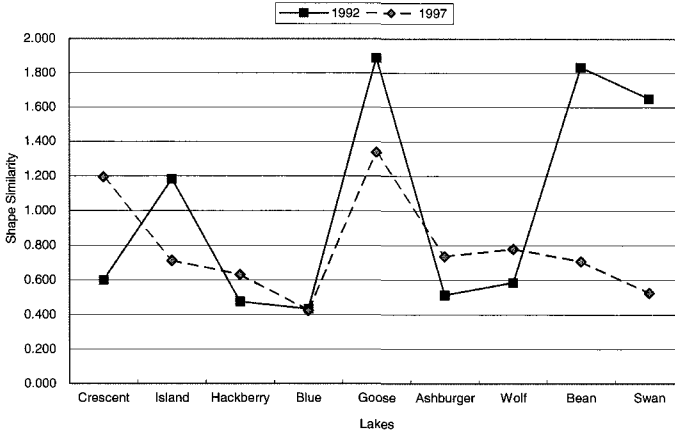


Fig. 18 (b). Shape similarity of the lakes in Autumn.

## 7. Conclusions

In this chapter, we have presented a shape-based approach to lake change detection using time series Landsat images. Using support vector machine, we achieved higher classification accuracy than that using other traditional classifiers. The study results show the effectiveness of the proposed shape retrieval and representation scheme and its potential capability in specific applications in the study of lake change response to short or long term climatic variation and drought monitoring. The correlation between the lake area and the drought indices is regarded as significant, although it is beyond the scope of this paper to discuss detailed reasons for it. Shape similarity is also considered important in quantitative measurement of the object variation besides the classical parameters such as area, perimeter, etc. However, the limitation of the temporal resolution of the available image data sets prevents us from performing further detailed trend analysis. In addition, images with higher spatial resolution are preferred in order to characterize the shapes of the lakes more precisely. Finally, the proposed approach can be used to automatically retrieve and represent other sorts of objects with irregular shapes, although it was applied to a special kind of object of interest, namely, the lake, in this research.

## Acknowledgment

The authors are thankful to Professor Donald C. Rundquist at UNL for providing the image data sets. The authors also thank Dr. Joseph Mitchell at Cornell University for providing the source code for the shape similarity measure.

## References

- [1] J. A. Richards, *Remote Sensing Digital Image Analysis: An Introduction*, Springer-Verlag, Berlin, 1993, pp. 130-131.
- [2] F. Dell'Acqua and P. Gamba: "Query-by-shape in meteorological image archives using the Point Diffusion Technique", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 9, pp. 1834-1843, Sept. 2001.
- [3] R. C. Veltkamp and M. Hagedoorn, "State-of-the-art in shape matching," *Technical Report UU-CS-1999-27*, Utrecht University, Netherlands, 1999.
- [4] L. F. Costa, R.M. Cesar, *Shape Analysis and Classification: Theory and Practice*, CRC Press: Boca Raton, 2001.
- [5] A. Cumani, "Edge detection in multispectral images," *CVGIP: Graphical Models and Image Processing*, vol. 53, pp. 40-51, January 1991.
- [6] J. C. Tilton and W. T. Lawrence, "Interactive analysis of hierarchical image segmentation," in *Proceedings of the IEEE IGARSS 2000*, vol. 2, Honolulu, Hawaii, USA, July 2000, pp. 733-735.
- [7] J. Le Moigne and J. C. Tilton, "Refining image segmentation by Integration of Edge and Region Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, pp. 605-615, May 1995.
- [8] D. W., Buckwalter, "Monitoring Nebraska's Sandhills lakes", *Resource Report No. 10*, Conservation and Survey Division, Institute of Agriculture and Natural Resources, University of Nebraska-Lincoln, Lincoln, Nebraska, 1983.
- [9] D. C. Rundquist, M. P. Lawson, L. P. Queen, and R. S. Cerveny, "The relationship between summer-season rainfall events and lake-surface area", *Water Resources Bulletin*, vol. 23, no. 3, pp. 493-508, June 1987.
- [10] D. C. Gosselin, D. C. Rundquist, and S. K. McFeeters, "Remote monitoring of selected ground-water dominated lakes in the Nebraska Sand Hills," *Journal of American Water Resource Association*, vol. 36, no. 5, pp. 1039-1051, Oct. 2000.
- [11] E. A. Work and D. S. Gilmer, "Utilization of satellite data for inventorying prairie ponds and lakes," *Photogrammetric Engineering and Remote Sensing*, vol. 42, no. 5, pp. 685-694, 1976.
- [12] C. Cortes and V.N. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, pp. 1-25, 1995.
- [13] C. J.C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, 1998.
- [14] S. Fukuda and H. Hirokawa, "Support vector machine classification of land cover: application to polarimetric SAR data," in *Proceedings of the IGARSS 2001*, vol. 1, Sydney, Australia, July 2001, 187-189.
- [15] J. A. Gualtieri, S. R. Chettri, R. F. Cromp, and L. F. Johnson, "Support vector machine classifiers as applied to AVIRIS data", in R. Green, editor, *Summaries of the Eighth JPL Airborne Earth Science Workshop: JPL Publication 99-17*, NASA/JPL, Feb 1999, pp. 217-227.
- [16] M. R. Azimi-Sadjadi and S. A. Zekavat, "Cloud classification using support vector machines", in *Proceedings of the IEEE IGARSS 2000*, vol.2, Honolulu, Hawaii, USA, July 2000, pp. 669-671.
- [17] S. L. Tanimoto, *The Elements of Artificial Intelligence Using Common Lisp*, W. H. Freeman, New York, 1990.

- [18] A. Rosenfeld and J. L. Pfaltz, "Sequential operations in digital picture processing," *Journal of the Association for Computing Machinery*, vol. 13, pp. 471-494, 1996.
- [19] R. Lumia, L. G. Shapiro, and O. Zuniga, "A new connected components algorithm for virtual memory computers," *Computer Vision, Graphics, and Image Processing*, vol. 22, pp. 287-300, 1983.
- [20] L. G. Shapiro and G. C. Stockman, *Computer Vision*, Prentice Hall, New Jersey, 2001, pp. 56-63.
- [21] I. Tomek, "Two algorithms for piecewise-linear continuous approximation of functions of one variable," *IEEE Transactions on Computers*, vol. 23, no.4, pp. 445-448, 1974.
- [22] K. Wall and P.E. Danielsson, "A fast sequential method for polygonal approximation of digitized curves," *Computer Vision, Graphics Image Process*, vol. 28, pp. 220-227, 1984.
- [23] D. Magee and R. Boyle, "Building shape models from image sequences using piecewise linear approximation," In *Proc. British Machine Vision Conference*, Highfield, Southampton, UK, September, 1998, pp. 398-408.
- [24] R. Chen, M. Ouyang, and P. Revesz, "Approximation data in constraint databases," in *Proc. The 4th International Symposium on Abstraction, Reformulation and Approximation*, Horseshoe Bay, Texas, July 2000, pp. 124-143.
- [25] P. Revesz, *Introduction to Constraint Databases*, Springer: Telos, 2001, pp. 250-263.
- [26] E. M. Arkin, L. P. Chew, D. P. Huttenlocher, K. Kedem, and J. S. B. Mitchell, "An efficiently computable metric for comparing polygonal shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 3, pp. 209-216, March 1991.
- [27] J. R. Jensen, *Introductory Digital Image Processing: A Remote Sensing Perspective*, Prentice-Hall, New Jersey, 1996.
- [28] N. Cristianini, and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, The Cambridge University Press, Cambridge, UK, 2000.
- [29] C. Chang and C. Lin, "LIBSVM: a library for support vector machines," 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [30] T. Joachims, "Making large-scale SVM learning practical," in *Advances in Kernel Methods – Support Vector Learning*, MIT Press, 1999, pp. 169-184.
- [31] V. N. Vapnik, *Statistical Learning Theory*, John-Wiley and Sons, Inc, 1998, pp. 423-424.
- [32] B. Scholkopf, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing support vector machines with Gaussian kernels to Radial Basis Function classifiers," *Technical Report A. I. Memo No. 1599*, Massachusetts Institute of Technology, December, 1996.
- [33] S. D. Kim, J. H. Lee, and J. K. Kim, "A new chain-coding algorithm for binary images using run-length codes," *Computer Vision, Graphics and Image Processing*, vol. 41, pp. 114-124, 1988.
- [34] H. Freeman, "Computer processing of line-drawings images," *Computing surveys*, vol.6, no. 1, pp. 57-95, 1974.
- [35] W. M., Alley, "The Palmer Drought Severity Index: limitations and assumptions," *Journal of Applied Meteorology*, vol. 23, pp. 1100-1109, 1984.

# **SAR Image Processing**

This page is intentionally left blank



## CHAPTER 5

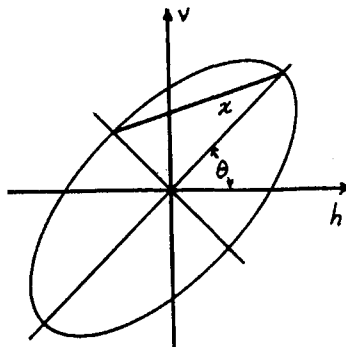
### A REVIEW OF POLARIZATION ORIENTATION ANGLE ESTIMATION FROM POLARIMETRIC SAR DATA

Jong-Sen. Lee, Dale L. Schuler and Thomas L. Ainsworth  
*Remote Sensing Division, Naval Research Laboratory, Washington Dc 20375-5351*  
*E-mail: lee@ccf.nrl.navy.mil*

This chapter reviews the recent advances in polarization orientation angle estimation and its applications using polarimetric synthetic aperture radar (SAR) data. Polarization orientation shifts are induced by topography slopes in the azimuth direction. Orientation angles can be readily extracted from polarimetric SAR data. Difficulties are frequently encountered in the estimation of orientation angles from polarimetric data. These difficulties will be discussed and the effect of radar wavelength and calibration on the estimation will be investigated. SIR-C and JPL AIRSAR polarimetric SAR data are used for illustration.

#### 1. Introduction

Interferometric (SAR) has been successfully applied to measure topography. Radar interferometry requires the use of dual antennas separated by a baseline in a single pass system, or by a repeated pass configuration. Recently, a new technique has been developed using polarimetric SAR (POLSAR) to measure azimuth slopes that are related to shifts in polarization orientation angles [1-4]. Polarization orientation angle is one of the least used parameters among the wealth of polarimetric information when analyzing POLSAR data. The polarization state of an electromagnetic wave is characterized by its polarization orientation angle  $\theta$  and ellipticity angle  $\chi$  as shown in Fig. 1. The orientation angle, which is of importance to this study, is the angle between the major axis of the polarization ellipse and the horizontal axis. For distributed media, orientation shifts are induced by azimuthal slopes, which cause the polarization rotation about the line of sight.

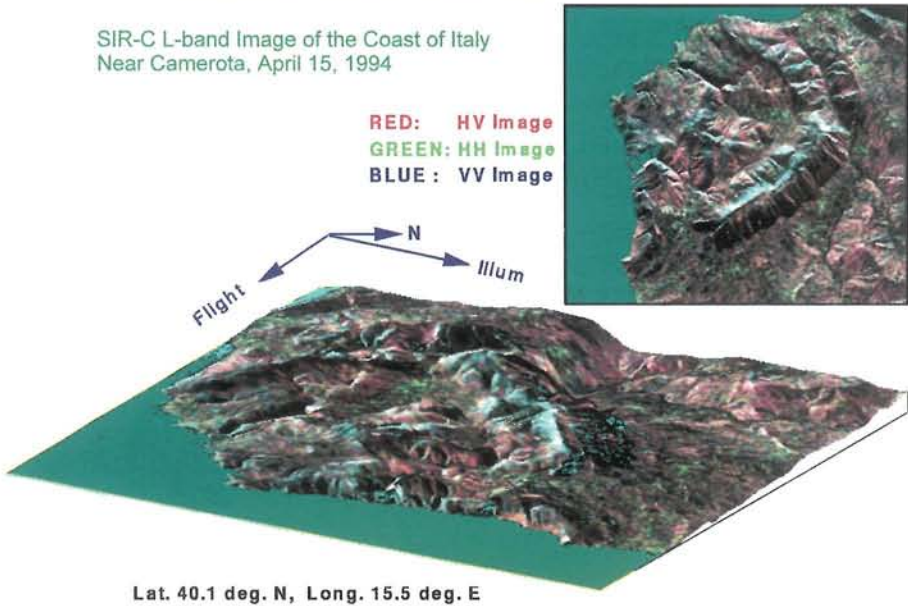


*Fig. 1. Polarization ellipse and the orientation angle.*

Polarization orientation shifts are frequently considered as a direct measure of azimuthal slopes. This is not correct. Lee [4] and Pottier [5] have found that

orientation shifts are also affected by the radar look angle and the range slope. However, in the absence of range slope data, orientation angles can be used as a rough estimate of azimuthal slopes, especially in cases of gentle terrain and a large radar look angle. An example is shown in Fig. 2. A SIR-C L-band polarimetric SAR data of Camerota, Italy, is used in the computation of orientation angles. The upper right image is the SAR image, color coded by  $|HV|$ ,  $|HH|$ , and  $|VV|$  as red, green and blue, respectively. Orientation angles were estimated from the polarimetric data. A DEM was generated by integrating orientation angles as azimuthal slopes. When integrating azimuthal slopes to produce a DEM, we use the coastline to provide a common zero elevation reference. The 3-D representation of the derived DEM, with the polarization SAR image (upper right) draped over it, is shown in the lower part of Fig. 2. It is evident that polarization orientation angles can be extracted from POLSAR data, and that they are related to the azimuthal slopes.

### POLARIMETRIC SAR DERIVED TOPOGRAPHY



*Fig. 2 Polarimetric SAR derived topography based on SIR-C L-Band POLSAR data. Orientation angles derived from the POLSAR data are used to produce a DEM.*

In this chapter, we review orientation angle estimation methods, and the radar geometry related to the azimuth and range slopes. Difficulties are frequently encountered in the estimation of orientation angles from POLSAR images. These difficulties will be discussed and the effect of radar wavelength and calibration on the estimation will be investigated. Applications to geophysical parameter estimation, and to ocean surface feature sensing will be specifically mentioned.

## 2. Radar Geometry of Polarization Orientation Angle

The change in the polarization orientation angle is geometrically related to topographical slopes and the radar look angle [3]. Fig. 3 shows the schematic diagram. The unit vector pair  $(\hat{x}, \hat{y})$  defines a horizontal plane,  $(\hat{y}, \hat{z})$  defines the radar incidence plane, and the radar line of sight is in the reverse direction of the axis  $\hat{I}_1$ . The angle  $\phi$  between  $\hat{I}_1$  and  $\hat{z}$  is the radar look angle. The axis  $\hat{x}$  is in the azimuth direction, and  $\hat{y}$  is in the ground range direction. The surface normal for a ground patch is denoted by  $\hat{N}$ . Assume that the polarimetric SAR is calibrated so that the horizontal polarization (H) is parallel to the horizontal plane  $(\hat{x}, \hat{y})$ , and the vertical polarization (V) is in the incidence plane.

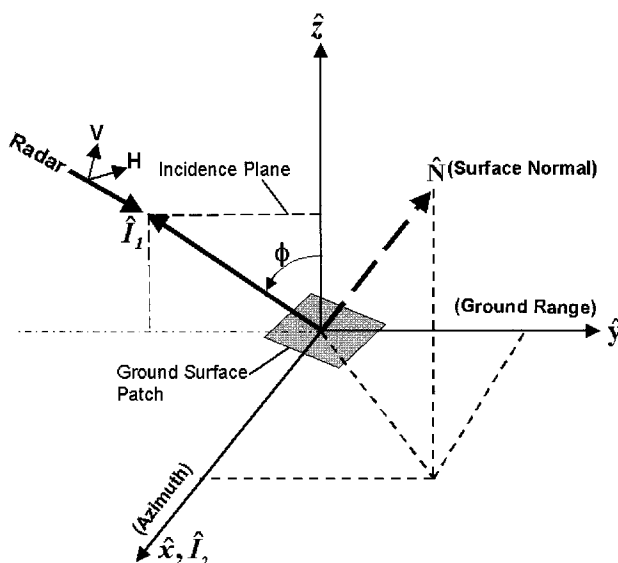


Fig. 3 A schematic diagram of the radar imaging geometry which relates the orientation angle to the ground slopes.

For a horizontal surface patch, its surface normal  $\hat{N}$  is in the incidence plane, and no orientation angle shift is induced. However, for a surface patch with an azimuthal tilt, its surface normal  $\hat{N}$  is no longer in the incidence plane. The induced polarization orientation angle shift  $\theta$  is the angle that rotates the incidence plane  $(\hat{y}, \hat{z})$  about the line of sight to the surface normal by the following equation [3],

$$\tan \theta = \frac{\tan \omega}{-\tan \gamma \cos \phi + \sin \phi} \quad (1)$$

where  $\tan\omega$  is the azimuth slope, and  $\tan\gamma$  is the slope in the ground range direction. This equation shows that the orientation angle shift is mainly induced by the azimuth slope, but that it is also a function of the range slope and the radar look angle. For small range slope, the orientation angle tends to overestimate azimuthal slope angle by the factor of  $(1/\sin\phi)$ . In general, orientation angle measurements overestimate the actual azimuth slope angles, when the range slope is positive (toward the radar), and may underestimate them, if the range slope is negative. The difference between the orientation angle and the corresponding azimuth slope angle becomes smaller for larger radar look angles. For an accurate estimate of azimuth slopes, range slope information, therefore, is required. This can be achieved by imaging the area with polarimetric SAR in orthogonal passes [2].

### 3. Polarimetric SAR Data Representation and Rotation

To derive the orientation angle estimation algorithm, it is necessary to understand the rotation of polarimetric matrices and the transformation to a circular polarization basis.

#### Scattering Matrix

POLSAR data can be represented by the scattering matrix (2) for single-look complex data.

$$\mathbf{S} = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \quad (2)$$

For backscattering from reciprocal media,  $S_{HV} = S_{VH}$ . The rotation of an orientation angle  $\theta$  is achieved by

$$\tilde{\mathbf{S}} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} S_{HH} & S_{HV} \\ S_{HV} & S_{VV} \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \quad (3)$$

The “ $\sim$ ” on top of the matrix  $\mathbf{S}$  denotes the matrix after the rotation by  $\theta$ . For convenience, this notation will be used throughout this chapter to indicate the matrix after the rotation by  $\theta$ .

#### Covariance Matrix from Circular Polarizations

The circular polarization components can be derived from the scattering matrix. The three circular components for right-right, left-left and right-left circular polarizations are

$$\begin{aligned} S_{RR} &= (S_{HH} - S_{VV} + i2S_{HV})/2 \\ S_{LL} &= (S_{VV} - S_{HH} + i2S_{HV})/2 \\ S_{RL} &= i(S_{HH} + S_{VV})/2 \end{aligned} \quad (4)$$

The rotation by an orientation angle can be obtained by applying (3). After manipulation, we have.

$$\begin{aligned}
\tilde{S}_{RR} &= S_{RR} e^{-i2\theta} \\
\tilde{S}_{LL} &= S_{LL} e^{i2\theta} \\
\tilde{S}_{RL} &= S_{RL}
\end{aligned} \tag{5}$$

Defining a circular basis vector,

$$c = \begin{bmatrix} S_{RR} \\ \sqrt{2}S_{RL} \\ S_{LL} \end{bmatrix} \tag{6}$$

the circular polarization based covariance matrix  $G$  is obtained from the vector  $c$  by

$$G = \langle c c^* T \rangle = \begin{bmatrix} \langle |S_{RR}|^2 \rangle & \sqrt{2} \langle (S_{RR} S_{RL}^*) \rangle & \langle (S_{RR} S_{LL}^*) \rangle \\ \sqrt{2} \langle (S_{RL} S_{RR}^*) \rangle & 2 \langle |S_{RL}|^2 \rangle & \sqrt{2} \langle (S_{RL} S_{LL}^*) \rangle \\ \langle (S_{LL} S_{RR}^*) \rangle & \sqrt{2} \langle (S_{LL} S_{RL}^*) \rangle & \langle |S_{LL}|^2 \rangle \end{bmatrix} \tag{7}$$

Applying (5), the upper off-diagonal terms of the rotated circular polarization covariance matrix are modified to (8).

$$\tilde{G} = \begin{bmatrix} \langle |S_{RR}|^2 \rangle & \sqrt{2} \langle (S_{RR} S_{RL}^*) e^{-i2\theta} \rangle & \langle (S_{RR} S_{LL}^*) e^{-i4\theta} \rangle \\ \sqrt{2} \langle (S_{RL} S_{RR}^*) e^{i2\theta} \rangle & 2 \langle |S_{RL}|^2 \rangle & \sqrt{2} \langle (S_{RL} S_{LL}^*) e^{-i2\theta} \rangle \\ \langle (S_{LL} S_{RR}^*) e^{i4\theta} \rangle & \sqrt{2} \langle (S_{LL} S_{RL}^*) e^{i2\theta} \rangle & \langle |S_{LL}|^2 \rangle \end{bmatrix} \tag{8}$$

The lower off-diagonal terms are the conjugate of their respective upper off-diagonal terms in (8). The diagonal terms are rotation invariant. It is apparent in (8) that the change in the orientation angle affects only the phases of off-diagonal terms. The circular polarization method to be discussed later is directly related to the (1, 3) term.

#### 4. The Circular Polarization Algorithm

The orientation angle shift causes rotation of both the scattering matrix and circular covariance matrix about the line of sight. Since the orientation angle information is embedded in the polarimetric SAR data, several methods have been developed to estimate azimuth slope induced orientation angles. The polarization signature method [1] and the circular polarization method [3] have been proven to be effective. Other methods have also been proposed [5-8]. The polarization signature method is based on the concept that the angle  $\theta$  corresponds to the change in the polarization orientation angle, and is estimated by the shift of the maximum co-polarization response. The polarization signature, as proposed by van Zyl [9], gives the polarization responses in the orientation and ellipticity plane, which is used to find the maximum co-polarization response. To speed up the optimization process, a steepest ascent algorithm was also developed [1].

In this chapter, we will limit the discussion to the circular polarization method, because it is based on a theoretical derivation. This method is also simpler and more accurate than any other methods.

The circular polarization method was proposed in [3] to extract the orientation angle, using right-right (RR) and left-left (LL) circular polarizations from either the single-look complex, or from the multi-look, data. This algorithm has proven to be successful. The concept of reflection symmetry [10] is used to explain the soundness of the circular polarization method, and to show problems associated with other algorithms. L-band POLSAR data of Camp Roberts, California are used to substantiate the value of this method.

Krogager and Czyz [6] proposed estimating orientation angles using the phase difference between right-hand and left-hand circular polarizations. This method has been further modified and refined by Lee et al. [3].

From the circular covariance matrix (8), the right-right and left-left circular polarization term (i.e., the (1,3) term) can be used to estimate the orientation angle. If the dominant orientation angle shift is induced by azimuth slope, then

$$\langle \tilde{S}_{RR} \tilde{S}_{LL}^* \rangle = \langle S_{RR} S_{LL}^* \rangle e^{-i4\theta} \quad (9)$$

For a reflection symmetrical medium as associated with a horizontal surface,  $\langle S_{RR} S_{LL}^* \rangle$  is required to be real in value, so that it will not corrupt the orientation angle related to the phase term  $e^{-i4\theta}$ . We will prove that  $\langle S_{RR} S_{LL}^* \rangle$  is real for a reflection symmetrical medium. For a reflection symmetrical medium, the cross-pol and co-pol correlation terms are zero. Substituting (4) into  $\langle S_{RR} S_{LL}^* \rangle$  and setting terms that contain  $S_{HV}$  to zero (except the  $\langle |S_{HV}|^2 \rangle$  term), we have

$$\langle S_{RR} S_{LL}^* \rangle = (-\langle |S_{HH} - S_{VV}|^2 \rangle + 4\langle |S_{HV}|^2 \rangle) / 4 \quad (10)$$

This term is real, so the argument of  $\langle S_{RR} S_{LL}^* \rangle$  is zero or  $\pi$ . Consequently, the phase difference between  $S_{HH}$  and  $S_{VV}$  does not cause error in the estimation of orientation angles.

The factor of  $4\theta$  in (9) limits the range of  $\theta$  to  $[-\pi/4, \pi/4]$ . To derive a general expression, substitute (4) into  $\langle \tilde{S}_{RR} \tilde{S}_{LL}^* \rangle$ . We then have

$$\langle \tilde{S}_{RR} \tilde{S}_{LL}^* \rangle = \frac{1}{4} \{ \langle -|\tilde{S}_{HH} - \tilde{S}_{VV}|^2 + 4|\tilde{S}_{HV}|^2 \rangle - i4 \text{Re}(\langle (\tilde{S}_{HH} - \tilde{S}_{VV}) \tilde{S}_{HV}^* \rangle) \} \quad (11)$$

From (9) and (11), we would have

$$-4\theta = \text{Arg}(\langle \tilde{S}_{RR} \tilde{S}_{LL}^* \rangle) = \tan^{-1} \left( \frac{-4 \text{Re}(\langle (\tilde{S}_{HH} - \tilde{S}_{VV}) \tilde{S}_{HV}^* \rangle)}{-\langle |\tilde{S}_{HH} - \tilde{S}_{VV}|^2 \rangle + 4\langle |\tilde{S}_{HV}|^2 \rangle} \right) \quad (12)$$

If (12) is applied directly, it would introduce errors, because, for an azimuth symmetrical medium,  $(\langle |\tilde{S}_{HH} - \tilde{S}_{VV}|^2 \rangle)$  is normally greater than  $4\langle |\tilde{S}_{HV}|^2 \rangle$ .

The denominator is then negative. Consequently, when the numerator is near zero, the arctangent is near  $\pm\pi$ . The orientation angle would be  $\pm\pi/4$  rather than near zero as it should be. To match the orientation angle corresponding to the azimuth slope angle, the bias must be removed by adding  $\pi$ . The circular polarization estimator is

$$\theta = \begin{cases} \eta, & \text{if } \eta \leq \pi/4 \\ \eta - \pi/2, & \text{if } \eta > \pi/4 \end{cases} \quad (13)$$

where

$$\eta = \frac{1}{4} \left[ \tan^{-1} \left( \frac{-4 \operatorname{Re}(\langle \tilde{S}_{HH} - \tilde{S}_{VV} \rangle \tilde{S}_{HV}^*)}{-|\tilde{S}_{HH} - \tilde{S}_{VV}|^2 + 4|\tilde{S}_{HV}|^2} \right) + \pi \right] \quad (14)$$

The arctangent in (14) is computed in the range of  $(-\pi, \pi)$ .



Fig. 4 This photo shows the topography and vegetation in Camp Roberts, California.

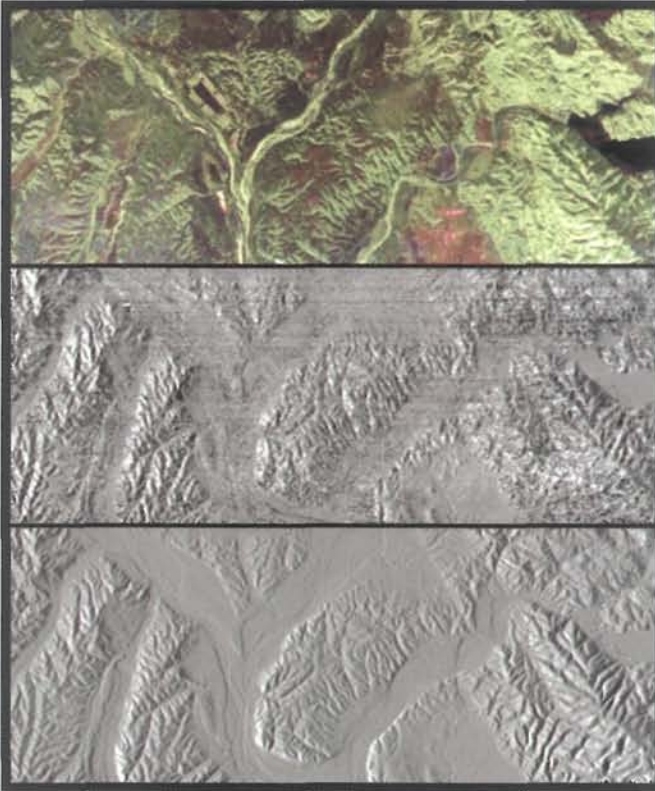
This algorithm has proven successful for orientation angle estimation [3]. An example is given here of applying it to the JPL AIRSAR L-band data of Camp Roberts, California. A photo of Camp Roberts in Fig. 4 shows the rugged terrain in the background with sparsely distributed oak trees. In the valley, the vegetation is much more dense.

The polarization image of Camp Roberts is shown in the top of Fig. 5. We use the Pauli matrix based color-coding for the combination of polarization channels: red for  $|\text{HH-VV}|$ , green for  $|\text{HV}|$ , and blue for  $|\text{HH+VV}|$ . The rectangular shaped object in the fork-like valley is the site of Camp Roberts. The middle image shows polarization orientation angles derived by the circular polarization method from the polarimetric data. The streaks at the top are from instrument noise.

JPL AIRSAR simultaneously imaged this area with C-Band TOPSAR to obtain interferometric data. This permits verification of polarimetric SAR derived orientation angles by those obtained from the interferometric generated



DEM and equation (1). Orientation angles derived from the DEM are shown in the lower image. The similarity between these two images indicates the validity of this estimation algorithm. The capability of deriving polarization orientation angles enables us to measure azimuthal slopes and to compensate polarimetric SAR data for terrain slope variation. The compensated data improves the accuracy of geophysical parameter estimations, as well as land-use and terrain type classification.



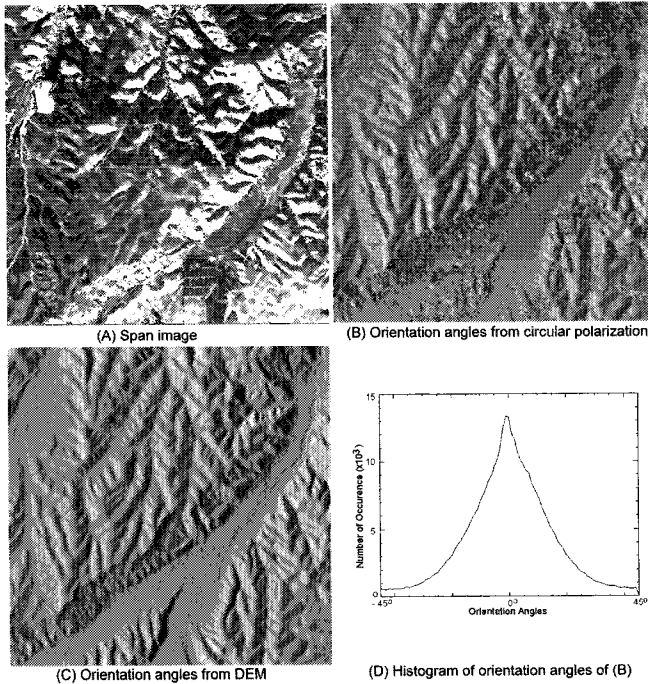
*Fig. 5 The top image shows the POLSAR data of Camp Roberts, The middle image shows polarization orientation angles derived by the circular polarization method. For comparison, the lower image shows orientation angles derived from a DEM, generated by C-band interferometric SAR. These two images are strikingly similar, except for the streaking in the middle image due to instrument noise.*

To take a closer look, an area within this image is selected which contains a variety of complex scatterers. Fig. 6A shows the span image of the selected area. The image size is 600x600 pixels. Rugged mountain terrain and a valley are present within the image. This POLSAR image contains artifacts, which appear as bright horizontal streaks. The orientation angle image derived by the circular polarization method is shown in Fig. 6B. For comparison, we computed the orientation angles from the interferometry generated DEM (shown in Fig. 6C). The circular polarization derived orientation angles show good agreement



with those derived from DEM. However, noisy results are scattered throughout the areas that correspond to bright areas in Fig. 6A. We observed that these areas also represent steep positive range slope areas that produce higher radar returns. For steep positive range slope areas, the scattering approaches the specular case, where  $\tilde{S}_{HH} \approx \tilde{S}_{VV}$ . In this situation, the measurement sensitivity is low for azimuth slope induced orientation angles. Consequently, the near specular scattering makes estimation very sensitive to vegetation variations.

A secondary effect is that high radar returns produce high speckle noise. Speckle noise is multiplicative in nature in the sense that the noise level is higher in higher return areas [11]. The bright specular returns from distributed media induce higher speckle noise in the orientation angle estimation. Fig. 6D shows the histogram of orientation angles produced by the circular polarization method. The bell shaped curve indicates that it is a good estimator compared with unsymmetrical histograms from most other methods.



*Fig. 6 Polarization orientation angles extracted from a 600x600 pixel area of Camp Roberts, California. (A) The L-band span image of an area containing a variety of complex scatterers. (B) The orientation angle image derived by the circular polarization method. (C) For comparison, the orientation angles from a DEM generated using C-band interferometry. (D) Histogram of orientation angles using the circular polarization algorithm.*

Comparisons of this algorithm with the original polarization signature algorithm and other methods based on polarimetric target decomposition [12] have been made by Lee et al.[3]. This algorithm is slightly better than the

polarization signature method, but is much better than those algorithms based on target decomposition.

## Remarks

### 1) Another estimator using the circular polarization

Another estimator can also be derived from the circular polarization matrix (8) using the (1, 2) term or the (2, 3) term. Assuming that the dominant orientation angle shift is induced by azimuth slopes, we have

$$\langle \tilde{S}_{RR} \tilde{S}_{RL}^* \rangle = \langle S_{RR} S_{RL}^* \rangle e^{-i2\theta} \quad (15)$$

If (15) is a valid estimator,  $\langle S_{RR} S_{RL}^* \rangle$  must not be complex in value for reflection symmetrical media. However, this assumption is not valid, because for reflection symmetrical media

$$\langle S_{RR} S_{RL}^* \rangle = [2 \langle \text{Im}(S_{HH} S_{VV}^*) \rangle - i(\langle |S_{HH}|^2 \rangle - \langle |S_{VV}|^2 \rangle)] / 4 \quad (16)$$

where  $\text{Im}(\ )$  denotes the imaginary part of a complex number. The above equation indicates that  $\langle S_{RR} S_{RL}^* \rangle$  is complex in value, except when  $S_{HH}$  and  $S_{VV}$  are in phase. For general terrain having vegetated or rough surfaces, the difference in the locations of scattering phase centers between HH and VV polarizations induces phase differences. Consequently, this estimator produces erroneous results in vegetated and rough surface areas.

### 2) Orientation Angle Estimation from the Scattering Matrix

Eq. (5) shows that the phase of  $\tilde{S}_{RR}$  is directly related to the orientation rotation. We assume  $S_{RR}$  to be real in value for a reflection symmetric medium. Applying (4) to  $\tilde{S}_{RR}$ , we have

$$|S_{RR}| \cos 2\theta + i |S_{RR}| \sin 2\theta = (\tilde{S}_{HH} - \tilde{S}_{VV} + i2\tilde{S}_{HV}) / 2 \quad (17)$$

If  $\tilde{S}_{HH}$ ,  $\tilde{S}_{VV}$  and  $\tilde{S}_{HV}$  are real in value, or in-phase, a simple estimator can be derived

$$\tan 2\theta = \frac{2\tilde{S}_{HV}}{\tilde{S}_{VV} - \tilde{S}_{HH}} \quad (18)$$

The above equation reveals that the orientation angle shift increases with  $\tilde{S}_{HV}$ . The range of the measurement from (18) appears to be between  $-\pi/2$  and  $\pi/2$ . The actual range is between  $-\pi/4$  and  $\pi/4$ , because the sign of the denominator is affected by the orientation rotation and scattering characteristics of the medium. In other words, the sign of the denominator is not uniquely affected by the orientation rotation. The algorithm is inherently unstable when  $\tilde{S}_{HH} \approx \tilde{S}_{VV}$ . This happens when radar scattering is near specular or when the orientation tilt nears  $45^\circ$ . In addition, scattering matrix data inherently has high speckle levels that make the result noisy for practical applications [11]. This algorithm is only of theoretical interest.

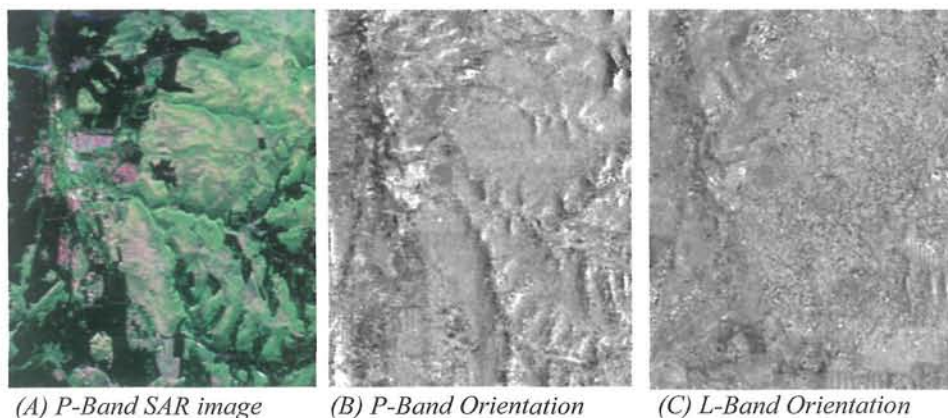
## 5. Discussion

Besides the selection of algorithms, many factors affect the accuracy of orientation angle estimations. They will be summarized in the following:

### Radar Frequency

Orientation angles can be derived from L-band and P-band POLSAR data, but less successfully from C-band or higher frequency data. Higher frequency POLSAR responses are less sensitive to azimuth slope variations, because electromagnetic waves with shorter wavelengths are less penetrative and are more sensitive to small scatterers within a resolution cell. The orientation angles induced from smaller scatterers overwhelm the orientation angle induced from the ground slope. We have found that C-band data produces a very noisy orientation angle image. On the other hand, longer wavelength radars (operating for example at P-band) are more penetrative and are less sensitive to smaller scatterers, and produce better results than L-band. Radio frequency interference, however, is often a problem.

JPL AIRSAR data from Freiburg Forest, Germany is used for illustration and is shown in Fig. 7. The area is heavily forested as shown in Fig 7A. The orientation angles derived from the P-band data (Fig. 7B) are well defined and show the strength of penetration from P-band. The orientation angles derived from the L-band data (Fig. 7C) are noisy, and are less sensitive to the under-canopy topography. C-band data produce even worse results than L-band.



*Fig. 7 In heavily forested areas, orientation angles can be extracted from P-band data, but not from L-band or higher frequency data. JPL AIRSAR P-band and L-band Data of Freiburg Forest, Germany, is applied to extract orientation angles. (A)  $|HH-VV|$ ,  $|HV|$  and  $|HH+VV|$  color coded P-Band SAR image, (B) Orientation angle image derived from the P-band data, (C) Orientation angle image derived from the L-band data.*

### The Importance of Polarimetric Calibration

POLSAR data calibration is a crucial step in the process of deriving accurate orientation angles. Both amplitude and phase calibration accuracies affect the derivation of orientation angles. The  $|S_{HV}|^2$  term and phase differences between cross-polarization and co-polarization terms are especially affected. Many polarimetric SAR calibration algorithms assume zero correlation between co-polarization and cross-polarization terms [12]. This assumption could introduce errors in orientation angle estimation. Recently, a revised method has been introduced by Ainsworth et al. [13] to account for this deficiency. In addition, non-zero pitch angles of the radar platform introduce a bias in the orientation angles. These pitch angles should be properly compensated before applying the orientation angle extraction method.

### Dynamic Range of Radar Response

The dynamic range and polarization channel isolation of the radar receiver are critical to the success of the orientation angle estimation. The success of the circular polarization methods depends on the accuracy of measuring the  $\langle (\tilde{S}_{HH} - \tilde{S}_{VV}) \tilde{S}_{HV}^* \rangle$  term. This term is much smaller than  $\langle |\tilde{S}_{HH}|^2 \rangle$  or  $\langle |\tilde{S}_{VV}|^2 \rangle$ . A lack of dynamic range makes this correlation term very noisy. In addition, POLSAR data compression, if necessary, has to be carefully devised to preserve the dynamic range. The extraction of orientation angles becomes an impossible task for SAR systems with small dynamic range and poor channel isolation.

## 6. Applications Involving Orientation Angles

### Polarimetric Data Compensation

The derived orientation angle can be used directly to compensate POLSAR data in rugged terrain areas. It is important to compensate the polarimetric SAR data to ensure accurate extraction of geophysical parameters, such as, soil moisture, surface roughness, snow cover, and biomass. A study on POLSAR data compensation has been investigated by Lee et al. [3]. The azimuth slope compensated data shows that all components of the coherency matrix have been modified except the  $\langle |\tilde{S}_{HH} + \tilde{S}_{VV}|^2 \rangle$  term, which is rotation invariant. The greatest changes occur in the real part of the  $\langle (\tilde{S}_{HH} - \tilde{S}_{VV}) \tilde{S}_{HV}^* \rangle$  term. The reduction in  $\langle |\tilde{S}_{HV}|^2 \rangle$  is also significant.

### DEM Generation

The derived orientation angles can be used to generate topography (Schuler et al. [1, 2]). Two orthogonal POLSAR flight passes are required to derive orientation angles in perpendicular directions. By applying equation (1), the ground slopes in two directions can be computed. The slope data is then used to solve a Poisson equation to estimate the elevation surface. This algorithm is similar to

the global least-square phase unwrapping algorithm used by SAR interferometry. Digital elevation maps have been generated. Due to the radar layover effect, difficulties were encountered when co-registering these two images. Currently, the accuracy of the DEM derived from this method is inferior to that generated by SAR interferometry.

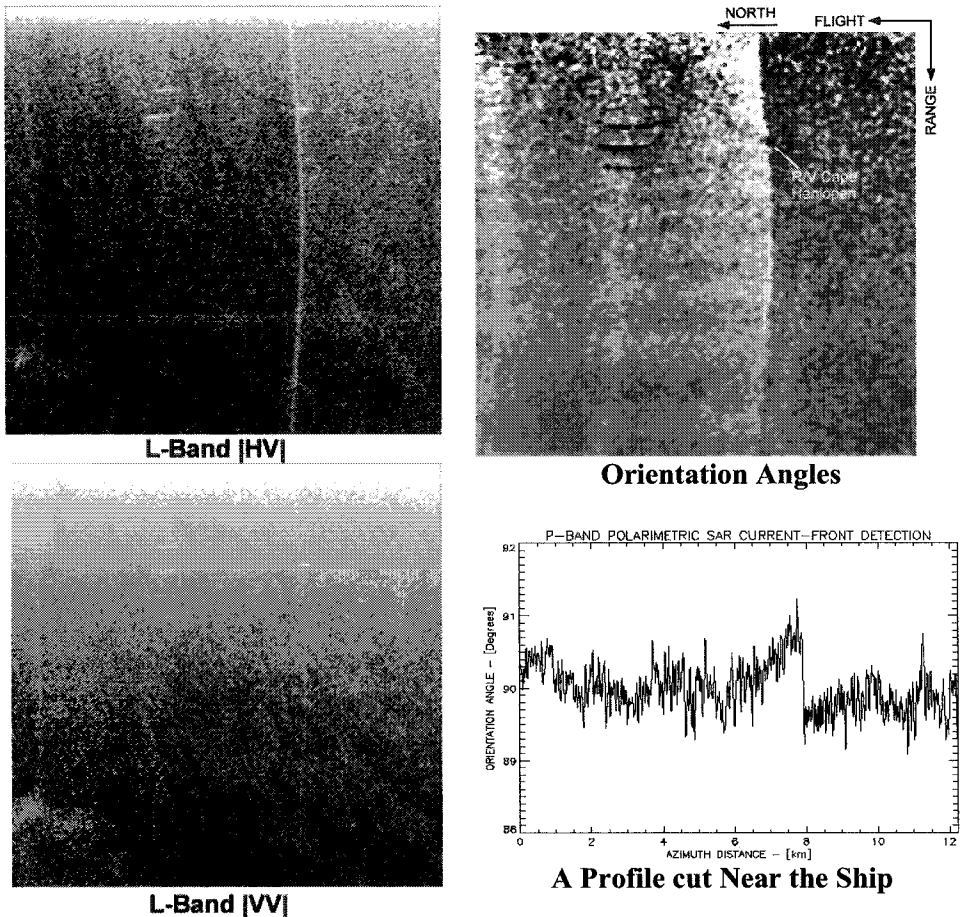


Fig. 8 Polarization orientation angles of a current front in the Gulf Stream. The  $|HV|$  and  $|VV|$  SAR images are shown on the left. The orientation angle derived from the P-band data is shown on the upper right, and a profile cut across near the ship is also shown.

### Ocean Applications

Another interesting application is for the direct estimation of ocean surface slopes. Backscattering from the ocean surface can be assumed in most cases to be homogeneous, and is characterized by two-scale Bragg scattering. This type of scattering provides excellent conditions for orientation angle estimation. Adjustment for range slopes according to (1) may not be necessary, because of

small range slopes of the ocean surface, but corrections for radar look angles have to be made.

In a study of convergent current fronts within the Gulf Stream (Lee et al. [14]), it was found that there existed a sudden change in the orientation angle from positive to negative across a convergent front with the maximum slope change being smaller than  $2^\circ$  as shown in the orientation angle image on the upper right of Fig. 8. The  $|HV|$  image (upper left of Fig. 8) shows the front as a bright linear feature, but the signature of the front is much weaker in the  $|VV|$  image. This study indicates the potential to use orientation angles to estimate small ocean surface slopes within an accuracy of a fraction of a degree. This study has been expanded by Schuler et al. [15, 16] and Kasilingam et al. [8] to estimate ocean wave slope spectra, and by Schuler et al [17] who applied this technique to study internal wave radar signatures. In addition, Ainsworth et al. [18] used this technique to study ocean surface features.

## 7. Conclusion

In this chapter, we have reviewed the recent advances in polarization orientation angle related research. Polarization orientation angle estimations based on the circular polarization covariance matrix are reviewed in detail. The concept of reflection symmetry was used to explain the soundness of the circular polarization method, and to show problems associated with other algorithms. Difficulties encountered in the estimation of orientation angles are discussed. We believe that this technique should be applied for better accuracy in geophysical parameter estimation by compensating polarimetric SAR data in areas of high topographic relief. The potential for DEM generation and ocean remote sensing was also discussed. SIR-C and JPL AIRSAR L-band and P-band POLSAR data were used for illustrations.

## Acknowledgment

The authors acknowledge the support of Office of Naval Research (ONR) under the 6.1 Research. We thank the JPL AIRSAR team for their sustained effort in providing valuable polarimetric SAR data.

## References

- [1] D.L. Schuler, J.S. Lee, G. De Grandi, "Measurement of Topography Using Polarimetric SAR Images," *IEEE Trans. on Geoscience and Remote Sensing*, no.5, pp. 1266-1277, 1996.
- [2] D.L. Schuler, J.S. Lee, T.L. Ainsworth, and M.R. Grunes, "Terrain Topography Measurement Using Multipass Polarimetric Synthetic Aperture Radar Data," *Radio Science*, vol. 35, no.3, 813-832, May-June 2000.

- [3] J.S. Lee, D.L. Schuler and T.L. Ainsworth, "Polarimetric SAR Data Compensation for Terrain Azimuth Slope Variation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 5, 2153-2163, September 2000.
- [4] J.S. Lee, et al., "On the Estimation of Radar Polarization Orientation Shifts Induced by Terrain Slopes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 1, 30-41, January 2002.
- [5] E. Pottier, D.L. Schuler, J.S. Lee and T.L. Ainsworth, "Estimation of the Terrain Surface Azimuthal/Range Slopes Using Polarimetric Decomposition of POLSAR Data," *Proceedings of IGARSS'99*, 2212-2214, July 1999.
- [6] E. Krogager and Z.H. Czyz, "Properties of the Sphere, Diplane, and Helix Decomposition," *Proceedings of the Third International Workshop on Radar Polarimetry*, IRESTE, University of Nantes, pp.106-114, April 1995.
- [7] E. Pottier, "Unsupervised Classification Scheme and Topography Derivation of POLSAR Data on the «H/A/ $\alpha$ » Polarimetric Decomposition Theorem," *Proceedings of the Fourth International Workshop on Radar Polarimetry*, 535-548, Nantes, France, July 1998.
- [8] D. Kasilingam, H. Chen, D. Schuler and J.S. Lee, "Ocean Surface Slope Spectra from Polarimetric SAR Images of the Ocean Surface," *Proceedings of International Geoscience and Remote Sensing Symposium 2000*, 1110-1112, Honolulu, Hawaii, July 2000.
- [9] J.J. van Zyl, H.A. Zebker and C. Elachi, "Imaging Radar Polarization Signatures: Theory and Observations," *Radio Science*, vol.22, no. 4, 529-543, 1987.
- [10] S.H. Yueh, R. Kwok, and S.V. Nghiem, "Polarimetric Scattering and Emission Properties of Targets with Reflection Symmetry," *Radio Science*, vol. 29, no.6, pp.1409-1420, November 1994.
- [11] J.S. Lee, "Speckle Suppression and Analysis for Synthetic Aperture Radar Images," *Optical Engineering*, vol. 25, no. 5, 636-643, 1986.
- [12] S. Quegan, "A Unified Algorithm for Phase and Cross-Talk Calibration for Radar Polarimeters," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 32, no.1, pp. 89-99, 1994.
- [13] T.L. Ainsworth and J.S. Lee, "A New Method for a Posteriori Polarimetric SAR Calibration." *Proceeding of IGARSS2001*, Sydney, Australia, 9-13 July 2001.
- [14] J.S. Lee, et al., "Polarimetric Analysis and Modeling of Multifrequency SAR Signatures from Gulf Stream Fronts," *IEEE Journal of Oceanic Engineering*, vol.23, no. 4, pp.322-332, October 1998-160, 1998.
- [15] D.L. Schuler and J.S. Lee, "A Microwave Technique of Improve the Measurement of Directional Ocean Wave Spectrs," *Int. J. of Remote Sensing*, vol. 16, no. 2, pp. 199-215, 1995.
- [16] D.L. Schuler, "Measurement of Ocean Wave Spectra Using Polarimetric AIRSAR Data," *The 4<sup>th</sup> Annual JPL AIRSAR Geoscience Workshop*, Arlington, VA. 1993.

- [17] D.L. Schuler et al., "Polarimetric SAR Measurements of Slope Distribution and Coherence Change due to Internal Waves and Current Fronts," *Proceedings of GARSS 2002*, Toronto, Canada, June 2002.
- [18] T.L. Ainsworth, J.S. Lee and D.L. Schuler, "Multi-Frequency Polarimetric SAR Data Analysis of Ocean Surface Features," *Proceedings of International Geoscience and Remote Sensing Symposium 2000*, Honolulu, Hawaii, July 2000.



## CHAPTER 6

# UNSUPERVISED CLASSIFICATION OF NATURAL SCENES FROM POLARIMETRIC INTERFEROMETRIC SAR DATA

L. Ferro-Famil<sup>\*</sup>, E. Pottier<sup>\*</sup> and J.S. Lee<sup>\*\*</sup>

<sup>\*</sup> *IETR Laboratory, UMR CNRS 6164, Radar Polarimetry Group  
University of Rennes 1, Campus de Beaulieu – Bâtiment 11D  
263 Avenue General Leclerc, CS 74205 35042 Rennes Cedex, France  
E-mail: [laurent.ferro-famil@univ-renne1.fr](mailto:laurent.ferro-famil@univ-renne1.fr), [eric.pottier@univ-rennes1.fr](mailto:eric.pottier@univ-rennes1.fr)*

<sup>\*\*</sup> *Remote Sensing Division, Naval Research Laboratory,  
Washington, DC 20375-5351, USA  
E-mail : [lee@ccf.navy.mil](mailto:lee@ccf.navy.mil)*

This paper introduces an approach to the classification and interpretation of SAR data using complementary polarimetric and interferometric information. Strictly polarimetric and polarimetric interferometric data are first analyzed and classified separately. An unsupervised polarimetric segmentation, based on multivariate Wishart statistics, is applied to one of the separate polarimetric datasets. The use of pertinent polarimetric indicators permits to give an interpretation of each resulting cluster polarimetric properties and to classify the observed scene into three canonical scattering types. The interpretation and the segmentation of an optimized interferometric coherency set leads to the discrimination of different natural media that cannot be achieved with polarimetric data only. Finally, each type of scattering mechanism is processed through an unsupervised statistical interferometric classification procedure merging results from separate studies. The resulting classes show an enhanced description and understanding of the scattering from the different natural media composing the observed scene.

### 1. Introduction

There is currently a great deal of interest in the use of polarimetry and interferometry for radar remote sensing. In this context, different and important objectives are to classify Earth terrain components within a fully polarimetric interferometric SAR image and then extract physical information from the observed scattering of microwaves by surface and volume structures.

Polarimetric SAR data classification has been widely addressed in the last decade<sup>1-6</sup>. The tight relation between natural media physical properties and their polarimetric features leads to highly descriptive classifications results that can be interpreted by analyzing underlying scattering mechanisms<sup>7-9</sup>. Interferometric data provide information concerning the coherence of the scattering mechanisms and

can be used to retrieve observed media structures and complexity<sup>10-12</sup>. An example of the complementary aspect of polarimetric and interferometric information is given with polarimetric interferometric data acquired by the DLR E-SAR sensor at L band in repeat-pass mode with a baseline of approximately 10m. Optical, polarimetric and interferometric images are displayed in Figure 1.

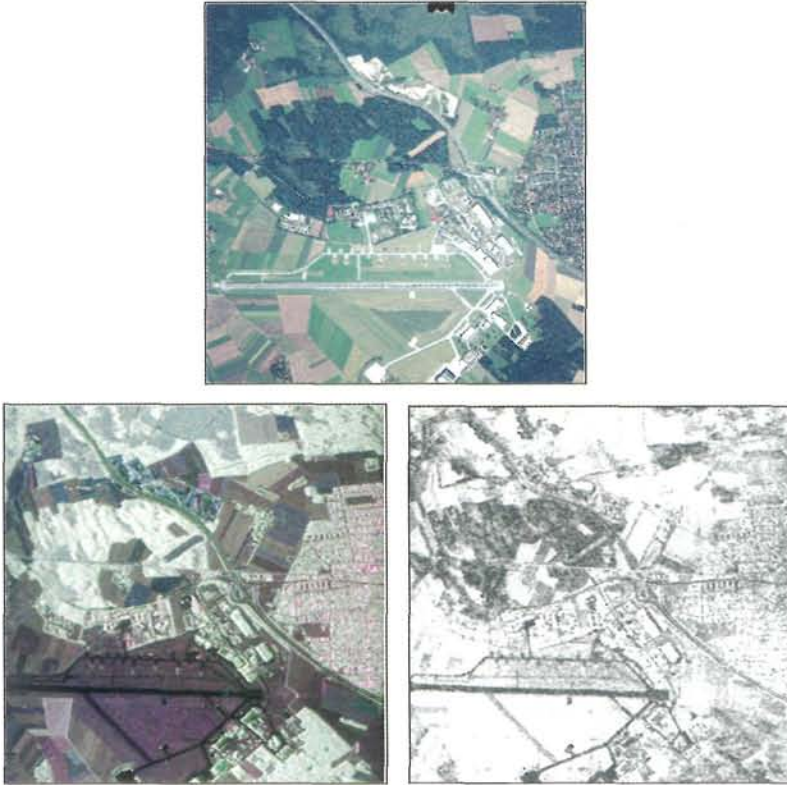


Figure 1 Optical image (top), polarimetric color coded image (bottom left) and interferometric coherence (bottom right) over the Oberpfaffenhofen test site

The Oberpfaffenhofen scene (Germany) is composed of various agricultural areas, forests and urban zones. Buildings at the center correspond to the DLR (German Aerospace Center). The polarimetric color-coded image and a coherence image give different descriptions of the observed scene. It can be observed in Figure 1 that, in general, forests have a uniform polarimetric behavior while the interferometric coherence show large variations. On the opposite some surface have similarly high coherences while the polarimetric image depicts different scattering mechanisms. The objective of an unsupervised classification process is to gather the complementary information contained in polarimetric and interferometric data

in order to deliver highly descriptive clusters as well as an interpretation of their characteristics.

The second section of this chapter presents an analysis of the characteristics of polarimetric SAR data based on the work of Cloude and Pottier<sup>7,8</sup>. A polarimetric decomposition theorem is applied to coherency matrices measured over natural targets in order to extract a dominant scattering mechanism. Parameters obtained from this decomposition theorem are interpreted in order to perform an unsupervised identification of the scattering mechanism.

The third section is dedicated to the unsupervised classification of polarimetric data based on the multivariate Wishart density function of coherency matrices. Results of the scattering mechanism identification procedure presented in part two are used to initialize two statistical segmentation schemes<sup>13-14</sup>. A k-mean clustering algorithm using Wishart statistics segments, in an unsupervised way, pixels into compact clusters according to their scattering properties. An identification of the resulting clusters to three canonical scattering mechanisms is introduced. It requires to estimate the number of relevant scattering contributions in order to reject secondary terms, which may introduce a bias in the retrieval of the scattering characteristics. Each scattering phenomenon is then identified to Volume diffusion, Surface Reflection or Double bounce Reflection<sup>15</sup>.

In the fourth part is presented the analysis of polarimetric interferometric data from their (6×6) coherency matrix. The optimization procedure developed by Papathanassiou and Cloude<sup>10-12</sup> is used to process optimal coherence sets showing an enhanced contrast and being highly related to the target volume structure.

This optimal coherence set is normalized in part five and parameterized with two indicators<sup>15</sup>. These parameters are then used to segment in an unsupervised way natural media with different coherent features. Finally results from the polarimetric identification procedure and from the analysis of the optimal interferometric coherence spectrum are merged into a classification procedure based on the Wishart statistics of the interferometric coherency matrix<sup>6,16</sup>.

The effectiveness of this classification method is demonstrated on polarimetric interferometric SAR data acquired by the DLR E-SAR sensor at L band over the Oberpfaffenhofen test site (Germany).

## 2. Polarimetric SAR Data Analysis

### 2.1. *Incoherent Polarimetric Representation*

For a given measurement configuration, a target is fully characterized by its coherent (2×2) complex scattering matrix,  $S$ , relating the incident and scattered Jones vectors. In a monostatic case, this matrix is composed, in the (X,Y) polarization basis, of three complex variables and is given by :

$$\mathbf{S} = \begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix} \quad (1)$$

An important development in the extraction of physical information from the scattering matrix has been achieved through the construction of target vectors<sup>7,17</sup>. A three complex element target vector,  $\mathbf{k}$ , is built from a vectorization of the  $\mathbf{S}$  matrix as follows:

$$\mathbf{k} = [k_1, k_2, k_3]^T \quad \text{with } k_i = \frac{1}{2} \text{tr}(\mathbf{S}\Psi_i) \quad (2)$$

where  $\text{tr}(\mathbf{S}\Psi_i)$  stands for the trace of  $\mathbf{S}\Psi_i$  and  $^T$  indicates transposition.  $\Psi_i$  is a (2×2) matrix and belongs to the complex basis matrix set,  $\Psi$ , onto which the scattering matrix is decomposed. There exist, in the literature, many different basis sets. A special set, widely used in natural media remote sensing, is built from a linear combination of Pauli orthogonal matrix set elements and is given by:

$$\Psi = \left\{ \sqrt{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \sqrt{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \sqrt{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\} \quad (3)$$

The corresponding target vector,  $\mathbf{k}$ , has the following expression:

$$\mathbf{k} = \frac{1}{\sqrt{2}} [S_{xx} + S_{yy} \quad S_{xx} - S_{yy} \quad 2S_{xy}]^T \quad (4)$$

The factor  $\sqrt{2}$  arises in (3) and (4) to keep the norm of the target vector  $\mathbf{k}$  equal to the span of the scattering matrix  $\mathbf{S}$ , namely the total power back-scattered by the target, given by  $\text{span} = |S_{xx}|^2 + |S_{yy}|^2 + 2|S_{xy}|^2$ .

This type of target vector is generally used, in a linear basis, to retrieve natural media physical parameters. The amplitude of its elements is related to both the amplitude of the scattering matrix elements and to their relative phase difference. Data sets are generally processed using incoherent averaging techniques of data compression and/or speckle reduction. In order to characterize the second order statistical properties of a set of  $n$  independent  $\mathbf{S}$  matrices, a  $n$ -look (3×3) coherency matrix,  $\mathbf{T}$ , is built from the  $n$  corresponding target vectors  $\mathbf{k}_i$  as follows:

$$\mathbf{T} = \frac{1}{n} \sum_{i=1}^n \mathbf{k}_i \mathbf{k}_i^\dagger \quad (5)$$

where the symbol  $\dagger$  stands for transpose conjugate.

The diagonal terms of  $\mathbf{T}$  correspond to the back-scattered power in the different polarization channels while the off-diagonal terms relate to polarimetric correlation coefficients.

## 2.2. *H-A- $\alpha$ Polarimetric Decomposition Theorem*

An unfiltered coherency matrix is constructed from a single target vector and has its rank equal to one. Its elements are linked by four equations, known as "target equations"<sup>18</sup>, deriving from the calculation of the matrix minors. In this case, the coherency matrix is said to represent a pure target and is equivalent to a single relative scattering matrix,  $\mathbf{S}_{rel}$ , defined as:

$$\mathbf{S} = e^{j\theta_{rel}} \mathbf{S}_{rel} \quad (6)$$

If the number of independent samples is superior to one, the rank of  $\mathbf{T}$  is superior to ones and the coherency matrix cannot be related to a single relative scattering matrix.  $\mathbf{T}$  is then associated to a distributed target.

Cloude and Pottier<sup>7,8</sup> proposed an interpretation of the scattering phenomenon using an eigenvalue decomposition of the coherency matrix :

$$\mathbf{T} = \mathbf{V} \mathbf{\Sigma} \mathbf{V}^\dagger \quad (7)$$

where  $\mathbf{\Sigma}$  represents a diagonal non-negative real eigenvalue matrix and  $\mathbf{V}$  is a special unitary eigenvector matrix.

The eigenvector decomposition of a distributed target coherency matrix is considered as a simple statistical model consisting in the expansion of  $\mathbf{T}$  into a weighted sum of three coherency matrices  $\mathbf{T}_i$ . This expansion may be expressed as follows:

$$\mathbf{T} = \sum_{i=1}^3 \lambda_i \mathbf{v}_i \mathbf{v}_i^\dagger = \sum_{i=1}^3 \lambda_i \mathbf{T}_i \quad (8)$$

Unit power  $\mathbf{T}_i$  matrices are defined from the mutually orthogonal eigenvectors  $\mathbf{v}_i$ , while the expansion weights are given by the corresponding eigenvalues  $\lambda_i$ . Each coherency matrix,  $\mathbf{T}_i$ , can be related to a single relative scattering matrix.

If only one eigenvalue is non-zero,  $\mathbf{T}$  corresponds to a pure target and can be related to a single relative scattering matrix. On the other hand, if all eigenvalues are equal,  $\mathbf{T}$  is composed of three orthogonal scattering mechanisms with equal amplitudes. The corresponding target response shows no polarization structure at all. In general,  $\mathbf{T}$  has non-zero and non-equal eigenvalues. The analysis of its polarimetric properties requires a study of the eigenvalues distribution as well as a characterization of each scattering mechanism of the expansion.

### 2.2.1. Characterization of the eigenvalue set

Pseudo-probabilities of the coherency matrix expansion elements are defined, from a set of sorted eigenvalues, as:

$$p_i = \frac{\lambda_i}{\sum_{j=1}^3 \lambda_j} = \frac{\lambda_i}{span} \quad \text{with } \lambda_1 \geq \lambda_2 \geq \lambda_3 \quad (9)$$

The distribution of the three probabilities can be fully described by two parameters.

- The entropy,  $H$ , indicates the degree of statistical disorder of the scattering phenomenon. It derives, in the Von Neumann sense, from the set of probabilities as:

$$H = -\sum_{i=1}^3 p_i \log_3 p_i \quad \text{with} \quad 0 \leq H \leq 1 \quad (10)$$

If the entropy,  $H$ , is close to 0, the probability of the first term of the expansion described in (8) almost equals 1 and the coherency matrix  $\mathbf{T}$  represents a pure target.

Entropy close to 1 depicts a uniform distribution of the probability set. In this case the scattering process is a random noise and no reliable polarimetric information can be extracted from  $\mathbf{T}$ .

- For intermediate entropy values, i.e. different from 0 and 1, a complementary parameter is necessary to fully characterize the set of probabilities. The anisotropy,  $A$ , is defined as the relative importance of the secondary scattering mechanisms and is given by:

$$A = \frac{p_2 - p_3}{p_2 + p_3} \quad \text{with } 0 \leq A \leq 1 \quad (11)$$

A value of  $A$  close to 0 corresponds to secondary mechanisms with equal importance, while  $A=1$  indicates that the power associated to the third matrix,  $\mathbf{T}_3$ , is null.

### 2.2.2. Extraction of the dominant scattering mechanism

Cloude and Pottier<sup>7,8</sup> proposed a method to extract a pure dominant scattering mechanism from the results of the coherency matrix expansion enounced in (8). The unitary eigenvector associated to each scattering mechanism is parameterized, using four real coefficients, under the following form:

$$\mathbf{v}_i = e^{j\delta_i} [\cos \alpha_i, \sin \alpha_i \cos \beta_i e^{j\delta_i}, \sin \alpha_i \sin \beta_i e^{j\gamma_i}]^T \quad (12)$$

The phase term  $\xi_i$  depends on the way eigenvectors are calculated and does not contribute to the scattering mechanism. The coherency matrix expansion presented in (8) may then be rewritten as:

$$\mathbf{T} = \sum_{i=1}^3 \lambda_i \mathbf{T}_i(\alpha_i, \beta_i, \delta_i, \gamma_i) \quad (13)$$

The condition of mutual orthogonality between the eigenvectors involve that the three polarimetric parameters sets resulting from the expansion are not independent.

For this reason, each polarimetric parameter is associated to a three symbol Bernoulli statistical process. In this way, the estimate of the mean polarimetric parameter set is given by:

$$(\underline{\alpha}, \underline{\beta}, \underline{\delta}, \underline{\gamma}) = \sum_{i=1}^3 p_i (\alpha_i, \beta_i, \delta_i, \gamma_i) \quad (14)$$

The set of average polarimetric parameters defined in (14) may then be inserted in (11) to form an average target vector, leading to the definition of a pure coherency matrix corresponding to the dominant scattering mechanism. An interpretation of the polarimetric information related to each average parameter is given by Cloude and Pottier<sup>7</sup>.

- $\underline{\alpha}$  can be considered as the main parameter resulting from the estimation performed in (14). It indicates the nature of the underlying dominant scattering mechanism. A value close to 0 relates surface reflection,  $\underline{\alpha}$  equals 45° for scattering from a dipole and reaches 90° when the target consists in a metallic dihedral scatterer.
- The three remaining parameters were shown<sup>19</sup> to be related to the orientation of the considered target around the radar line-of-sight.

It may be easily shown that the indicator of the underlying scattering mechanism,  $\underline{\alpha}$ , is invariant under azimuthal rotation around the radar line-of-sight. Moreover, the entropy and anisotropy indicators, which are constructed from the set of eigenvalues, are invariant under any change of polarization basis. These invariance properties may confer to the three parameters an appreciable robustness for the characterization of general targets.

### 2.2.3. Application to polarimetric SAR data

The polarimetric decomposition described in the former paragraph is applied to the Oberpfaffenhofen polarimetric SAR dataset. The resulting entropy, anisotropy and  $\underline{\alpha}$  images are shown in Figure 2.

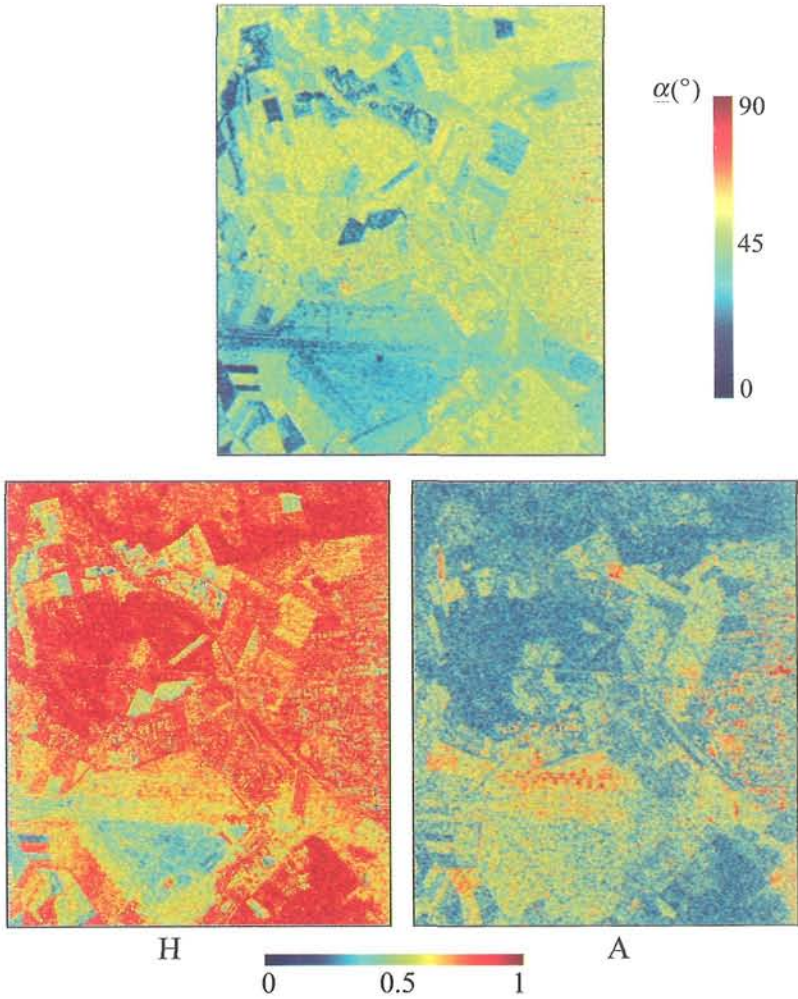


Figure 2 Polarimetric decomposition main parameters: H, A and  $\underline{\alpha}$

The values of  $\underline{\alpha}$  observed over the scene under analysis corroborate the interpretations presented in the former paragraph. In a general way,  $\underline{\alpha}$  is less than  $45^\circ$  degrees for a large part of the agricultural fields, which correspond to rough surfaces. It reaches  $40^\circ$  to  $60^\circ$  over forested areas, and shows values close to  $90^\circ$  for some buildings and urban areas characterized by double bounce scattering. The entropy has low values over surface scattering areas, which behave as almost deterministic targets. Over building and urban areas, H shows intermediate to high



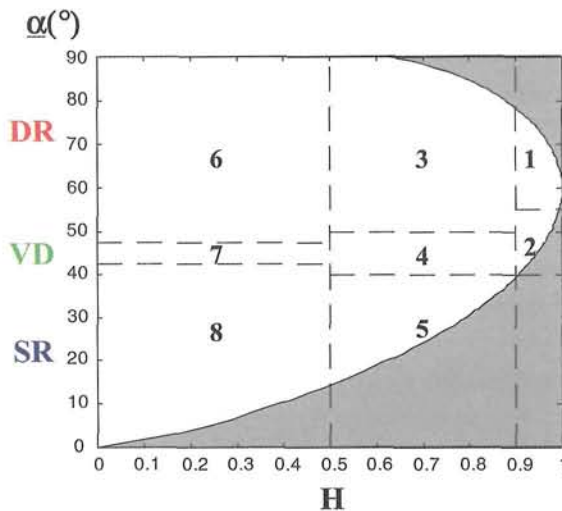
values due to the mixing of different types of scattering mechanism during the incoherent averaging. Some point targets, back-scattering fully polarized waves, are characterized by a null entropy. Over forests,  $H$  is close to 1 and depicts targets with no polarimetric structure.

### 2.3. Unsupervised Identification of a Polarimetric Scattering Mechanism

In <sup>7-8</sup>, an algorithm is proposed to identify in an unsupervised way polarimetric scattering mechanisms in the  $H$ - $\alpha$  plane. The key idea is that entropy arises as a natural measure of the inherent reversibility of the scattering data and that  $\alpha$  can be used to identify the underlying average scattering mechanism.

The  $H$ - $\alpha$  classification plane is sub-divided into 8 basic zones characteristic of different scattering behaviors. The basic scattering mechanism of each pixel of a polarimetric SAR image can then be identified by comparing its entropy and  $\alpha$  parameters to fixed thresholds. The different class boundaries, in the  $H$ - $\alpha$  plane, have been determined so as to discriminate surface reflection (SR), volume diffusion (VD) and double bounce reflection (DB) along the  $\alpha$  axis and low, medium and high degree of randomness along the entropy axis. Detailed explanations, examples and comments concerning the different classes can be found in <sup>7-8</sup>.

Figure 3 shows the  $H$ - $\alpha$  plane and the occurrence of the studied polarimetric data into this plane.



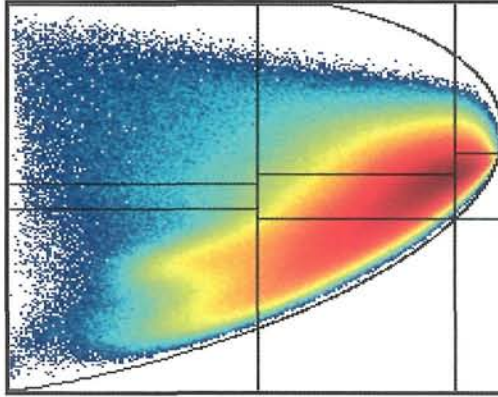


Figure 3 H- $\alpha$  scattering mechanism identification plane (top). Polarimetric data occurrence in the H- $\alpha$  plane (bottom)

Grey zones in the H- $\alpha$  plane correspond to unfeasible areas.

It can be seen, in Figure 3, that the largest densities in the occurrence plane correspond to volume diffusion and double bounce scattering with moderate and high randomness. Medium and low entropy surface scattering mechanisms are also frequently encountered in the scene under examination.

Data distribution in the H- $\alpha$  plane does not show, for the considered scene, distinct natural clusters belonging to a single scattering mechanism class. Therefore, identification results may highly depend on segmentation thresholds.

Results of the unsupervised identification procedure are presented in Figure 4.

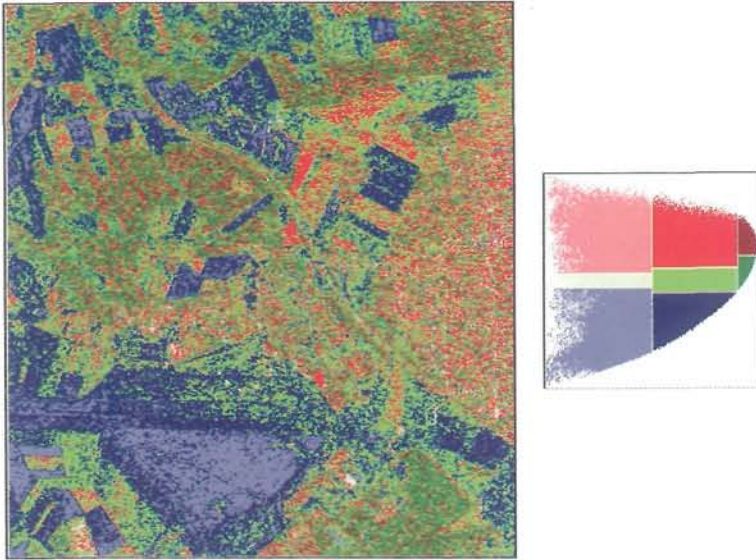


Figure 4 Unsupervised scattering mechanism identification in the  $H-\alpha$  plane

It can be observed in Figure 4 that the proposed segmentation in the  $H-\alpha$  plane permits to identify in a macroscopic way the type of scattering mechanism. Agricultural fields and bare soils are characterized by surface scattering. Scattering over forested areas is dominated by volume diffusion while urban areas are mainly characterized by double bounce scattering. It may be noted that the identification process slightly overestimates volume diffusion and double bounce scattering over surfaces.

### 3. Unsupervised Classification of Polarimetric SAR Data

The particularity of the identification procedure introduced in the former paragraph resides in the estimation of the type of observed media from a physical interpretation of canonical scattering mechanisms using robust indicators. Nevertheless, the analysis of natural scenes using this unsupervised approach may reach some limitations:

- The arbitrarily fixed linear decision boundaries in the  $H-\alpha$  plane may not fit data distribution. A natural cluster corresponding to similar targets may lie across a frontier in the decision plane. In this case, pixels with very similar characteristics may be assigned, in an almost random way, to different classes due slightly different locations in the  $H-\alpha$  plane. This effect can be observed in Figure 4 where the variability in natural media polarimetric features lead to noisy classification results.

- Even if the computation of  $H$  and  $\underline{\alpha}$  requires fully polarimetric data, these two parameters do not represent the whole polarimetric information. The use of other indicators such as the span or specific correlations coefficients may improve the classification results in a significant way.

Segmentation procedures based on the whole coherency matrix statistics permit to overcome the limitations mentioned above. Nevertheless, it is shown in the following, that the physical interpretation of the scattering phenomenon permits to enhance in a significant way the performance of statistical segmentation schemes.

### 3.1. Polarimetric SAR Data Statistics

It has been verified that when the radar illuminates an area of random surface of many elementary scatterers,  $\mathbf{k}$  can be modeled as having a multivariate complex gaussian probability density function  $N_c(\mathbf{0}, \Sigma)$  of the form<sup>4</sup>:

$$p(\mathbf{k}) = \frac{\exp(-\mathbf{k}^\dagger \Sigma^{-1} \mathbf{k})}{\pi^q |\Sigma|} \quad (15)$$

where  $q$  stands for the number of elements of  $\mathbf{k}$ , equal to three in the monostatic case,  $|\cdot|$  represents the determinant, and  $\Sigma = E(\mathbf{k}\mathbf{k}^\dagger)$  is the global (3×3) coherency matrix of  $\mathbf{k}$ .

It has been shown that assuming that target vectors have a  $N_c(\mathbf{0}, \Sigma)$  distribution, a sample  $n$ -look coherency matrix  $\mathbf{T}$ , introduced in (5), follows a complex Wishart distribution with  $n$  degrees of freedom,  $W_c(n, \Sigma)$ , given by<sup>4,20</sup>.

$$p(\mathbf{T}) = \frac{n^n |\mathbf{T}|^{n-q} \exp(-tr(n\Sigma^{-1}\mathbf{T}))}{K(n, q) |\Sigma|^n} \quad (16)$$

$$\text{with } K(n, q) = \pi^{q(q-1)/2} \prod_{i=1}^q \Gamma(n-i+1)$$

where  $\Gamma(\cdot)$  is the gamma function.

### 3.2. Unsupervised Maximum Likelihood Segmentation Based on the Wishart Distribution

A Maximum Likelihood (ML) segmentation process assigns sample coherency matrices  $\mathbf{T}$  to the class  $X_m$  maximizing its likelihood function over  $N$  possible classes. This decision may be expressed under the following form:

$$T \in X_m \text{ if } p(\mathbf{T}/X_m) > p(\mathbf{T}/X_i) \quad i = 1, \dots, N \quad i \neq m \quad (17)$$

The ML assignment of a sample coherency matrix following a Wishart distribution becomes:

$$X_m = \arg \max p(\mathbf{T} / X_m)$$

$$\text{with } p(\mathbf{T} / X_m) = \frac{n^q |\mathbf{T}|^{n-q} \exp(-tr(n\mathbf{\Sigma}_m^{-1}\mathbf{T}))}{K(n, q) |\mathbf{\Sigma}_m|^n} \quad (18)$$

where  $\mathbf{\Sigma}_m$  corresponds to the global coherency matrix evaluated over the class  $X_m$ .

The joint likelihood optimization for all the sample matrices cannot be performed in an easy way. Indeed, in the frame of an unsupervised segmentation, the global coherency  $\mathbf{\Sigma}_m$  is built from the sample matrices belonging to the class  $X_m$ . An optimal solution would consist in testing all the possible segmentations of a given number of sample matrices into  $N$  classes. This optimal solution cannot be applied due to the unrealistic computational load it requires.

Lee et al.<sup>4</sup> proposed an alternative method, based on a k-mean iterative clustering algorithm. At each iteration of the algorithm, a sample coherency is assigned to the class according to the following decision rule:

$$\mathbf{T} \in X_m \text{ if } d(\mathbf{T}, X_m) < d(\mathbf{T}, X_i) \quad i = 1, \dots, N \quad i \neq m \quad (19)$$

The statistical distance between the sample matrix  $\mathbf{T}$  and the class  $X_m$ ,  $d(\mathbf{T}, X_m)$ , derives from the Log-likelihood function and is given by:

$$d(\mathbf{T}, X_m) = \ln |\mathbf{\Sigma}_m| + tr(\mathbf{\Sigma}_m^{-1}\mathbf{T}) \quad (20)$$

Figure 5 depicts the unsupervised segmentation process.

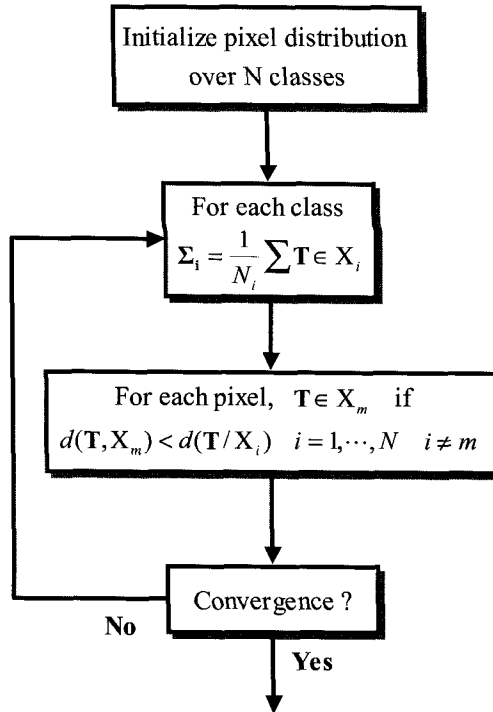


Figure 5 Unsupervised Maximum Likelihood segmentation scheme

It is known that the initialization of the pixel distribution into  $N$  classes is a critical stage of the  $k$ -mean clustering algorithm. An adequate initialization permits a fast convergence and provides correctly segmented clusters.

The convergence of the algorithm is evaluated by testing a condition of termination. Such a criterion may be defined from the estimation of the classification quality, or consist in a maximum number of iterations or in a sufficiently low number of pixels that are differently classified from one iteration to the other.

### 3.3. Combined Wishart $H$ - $A$ - $\alpha$ Unsupervised Classification Schemes

#### 3.3.1 Wishart $H$ - $\alpha$ unsupervised segmentation of polarimetric SAR data

Lee et al.<sup>13</sup> introduced an unsupervised ML segmentation procedure based on Wishart statistics, initialized with the results of the unsupervised identification of a scattering mechanism. This initialization provides 8 classes relating to the under-



lying physical scattering mechanism and giving a stable initial approximation of the segmentation.

Results obtained using Wishart  $H-\alpha$  segmentation are displayed in Figure 6.

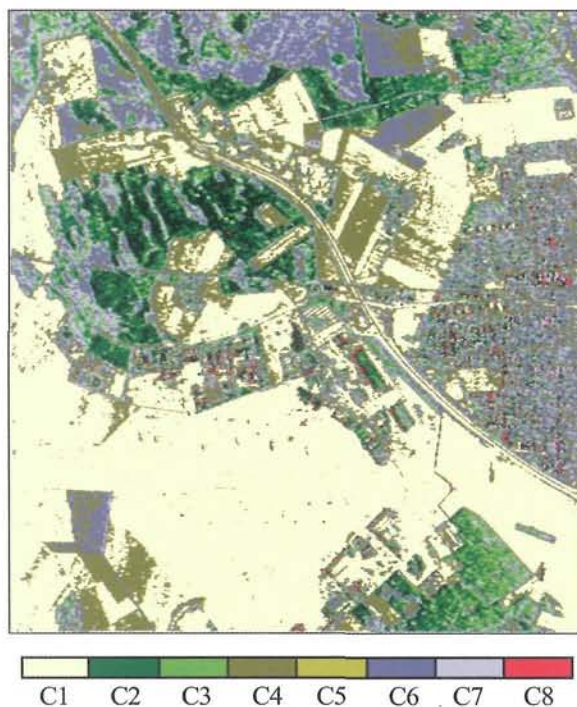


Figure 6 Wishart  $H-\alpha$  segmentation results

An important improvement in the segmentation accuracy can be observed in the image presented in Figure 6. The main kinds of natural media are clearly discriminated by the Wishart  $H-\alpha$  segmentation scheme. This unsupervised classification algorithm modifies the decision boundaries in an adaptive way to better fit the natural distribution of the scattering mechanisms and takes into account information related to the back-scattered power.

### 3.3.2 *Wishart $H-A-\alpha$ unsupervised segmentation of polarimetric SAR data*

Pottier and Lee<sup>14</sup> further improved the ML Wishart unsupervised segmentation by explicitly including the anisotropy information during the segmentation procedure. As mentioned previously, the anisotropy indicates the relative importance of sec-

ondary mechanisms obtained from the expansion of a coherency matrix. This polarimetric indicator is particularly useful to discriminate scattering mechanisms with different eigenvalue distributions but with similar intermediate entropy values. In such cases, a high anisotropy value indicates two dominant scattering mechanisms with equal probability and a less significant third mechanism, while a low anisotropy value corresponds to a dominant first scattering mechanism and two non-negligible secondary mechanisms with equal importance.

Among the different approaches tested, the best way to introduce the anisotropy information in the classification algorithm consists in implementing two successive segmentation procedures as shown in Figure 7.

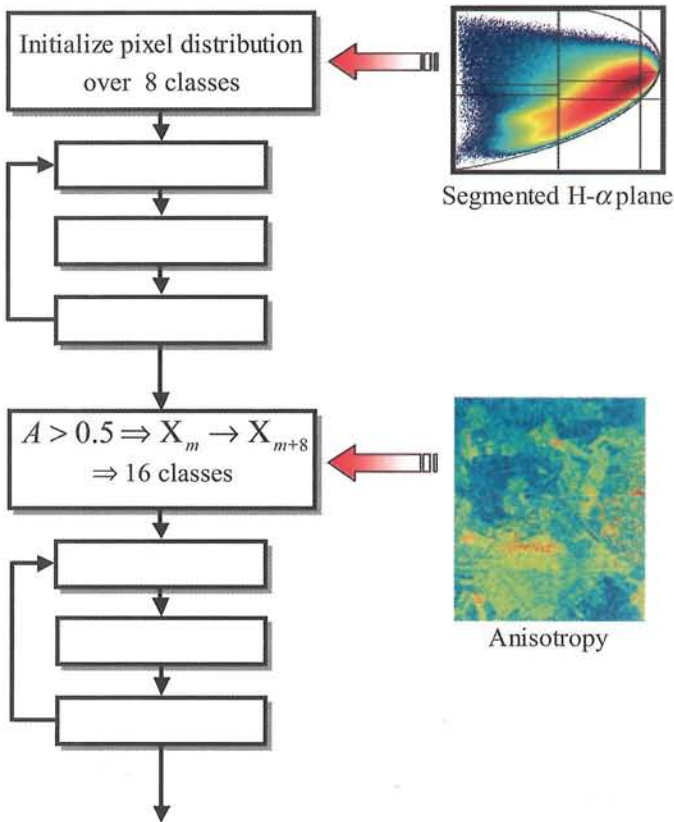


Figure 7 Synopsis of the Wishart H-A- $\alpha$  segmentation procedure



Polarimetric data are first segmented according to the algorithm presented in the former paragraph. Once this procedure has converged, the 8 resulting clusters are split into 16 ones by comparing the anisotropy of each pixel to a threshold fixed to 0.5. The 16 segments are then used to initialize a second Wishart ML segmentation procedure.

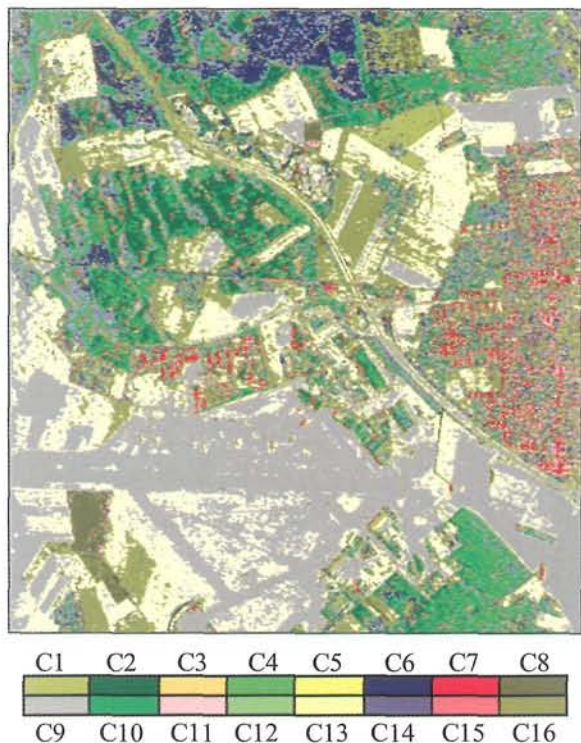


Figure 8 Wishart H-A- $\alpha$  segmentation results

The segmentation results presented in Figure 8 show an enhanced description of the Oberpfaffenhofen scene. The introduction of the anisotropy in the clustering process permits to split large segments into smaller clusters discriminating small disparities in a refined way. Several kinds of agricultural fields are separated. The runway and other low intensity targets are distinguished from other surfaces. Buildings are discriminated from other types of scatterers present in urban areas. The Wishart H-A- $\alpha$  classification scheme gathers into segments pixels with similar statistical properties, but does not provide any information concerning the nature of the scattering mechanism associated to each cluster.

### 3.4. Unsupervised Identification to Canonical Scattering Mechanisms

An efficient estimation of the nature of scattering mechanisms over natural scenes can be achieved by gathering results obtained from the polarimetric decomposition and segmentation procedures presented in previous paragraphs<sup>15</sup>. The identification of polarimetric properties of compactly segmented clusters permits, by analyzing groups of scatterers, to reduce the influence of the variations in polarimetric indicators over natural media. The estimation of scattering properties provides an accurate interpretation of nature media from manmade structure in the observed scene.

Volume diffusion and double bounce scattering were found to be over-estimated during the identification of scattering mechanisms using the H- $\alpha$  segmented plane. One of the reasons of this over-estimation resides in the calculation of the average parameter set given in (14). In some cases, the expansion of a coherency results in a dominant scattering mechanism and secondary mechanism showing very different polarimetric properties. The calculation of average indicators as the weighted sum of the indicators of each element of the expansion may then lead to an erroneous interpretation of the nature of the scattering mechanism. This phenomenon is illustrated by the following example:

$$\mathbf{T} = \begin{bmatrix} 0.4731 & -0.3242 & 0 \\ -0.3242 & 0.2369 & 0 \\ 0 & 0 & 0.29 \end{bmatrix} \quad (21)$$

$$\text{gives } \mathbf{V} = \begin{bmatrix} 0.8192 & 0 & 0.5735 \\ -0.5735 & 0 & 0.8192 \\ 0 & 1 & 0 \end{bmatrix} \text{ and } \mathbf{\Lambda} = \text{diag}[0.7, 0.29, 0.1]$$

The corresponding set of separate and global  $\alpha$  parameters is given by

$$[\alpha_1, \alpha_2, \alpha_3] = [35^\circ, 90^\circ, 55^\circ] \Rightarrow \underline{\alpha} = 51.15^\circ, H = 0.6, A = 0.93 \quad (22)$$

The example depicted in (21) and (22) shows that secondary mechanisms may corrupt average parameters and lead to the estimation of double bounce reflection, while the dominant contribution, with 70% of the total power, corresponds to surface scattering.

A restriction of the retrieval of the dominant scattering properties to the study of the most significant contributions permits to limit the influence of secondary mechanisms. As it was mentioned previously, the entropy and anisotropy parameters can be used to determine the number of relevant scattering mechanisms in the global scattering phenomenon.

In Figure 9 are represented different basic configurations of the pseudo-probability spectrum according to the values of H and A. This information is then

used to select, in the H-A plane, the appropriate contributions arising from the expansion of a coherency matrix<sup>15</sup>.

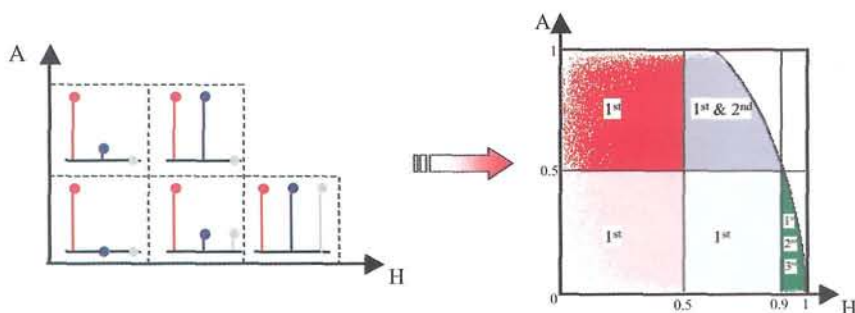


Figure 9 Discrimination of different pseudo-probability sets using H and A (left). Selection of relevant mechanisms in the H-A plane (right)

High and low entropy respectively correspond to random and almost deterministic scattering. Global scattering with intermediate entropy values are associated to two scattering mechanisms with equal importance or one dominant scattering mechanism perturbed by secondary terms according to the anisotropy value.

Specific identification procedures may then be applied to each of the three cases discriminated in the H-A plane according to the number of significant mechanisms<sup>15</sup>:

- In presence of a single dominant scattering mechanism, single and double bounce scatterings are separated by  $\alpha_i < 45^\circ$  and  $\alpha_i > 45^\circ$ , respectively.
- In the case of two significant mechanisms, scattering features cannot be estimated using average parameters in order to avoid the over-estimation problem mentioned above. Instead, a distributed matrix,  $\mathbf{T}_d$ , is constructed from the first two elements of the expansion as follows:

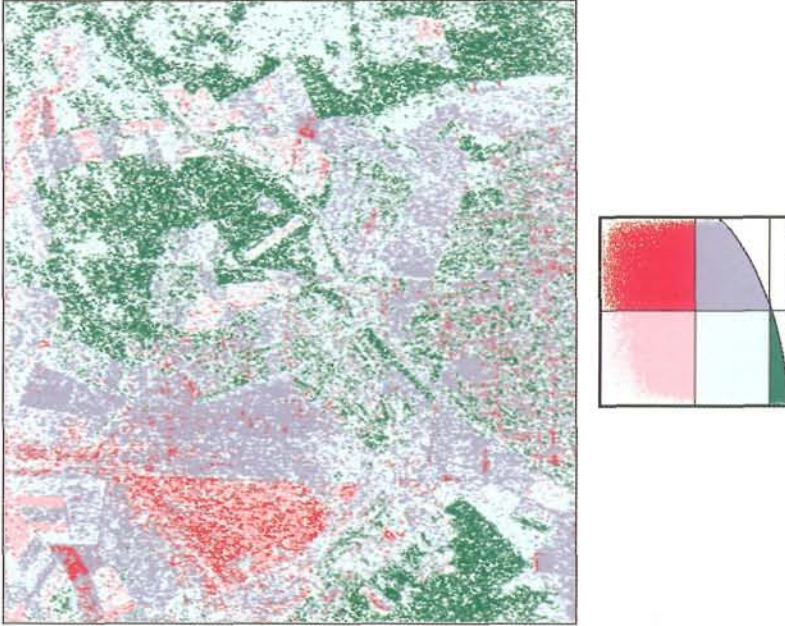
$$\mathbf{T}_d = p_1 \mathbf{T}_1 + p_2 \mathbf{T}_2 = \begin{bmatrix} 2A_0 & C - jD & H + jG \\ C + jD & B_0 + B & E + jF \\ H - jG & E - jF & B_0 - B \end{bmatrix} \quad (23)$$

where the distributed matrix is parameterized using the 9 Huynen coefficients. The nature of the scattering mechanism is determined by comparing the first two Huynen generators,  $2A_0$  and  $B_0 + B$ .

- Three mechanisms with equal probabilities depict random polarimetric scattering and are associated to volume diffusion.

The determination of the number of relevant mechanisms and the corresponding identification procedure are applied to each pixel of the image.

Its most representative scattering mechanism type then characterizes each cluster. It is important to note that this type of decision prevents an excessive sensitivity of the classification process to the hard-decision limits with respect to the parameters  $H$ ,  $A$  and  $\alpha$ . The relevant mechanisms selection and identification to canonical scattering mechanisms results are shown in Figure 10.





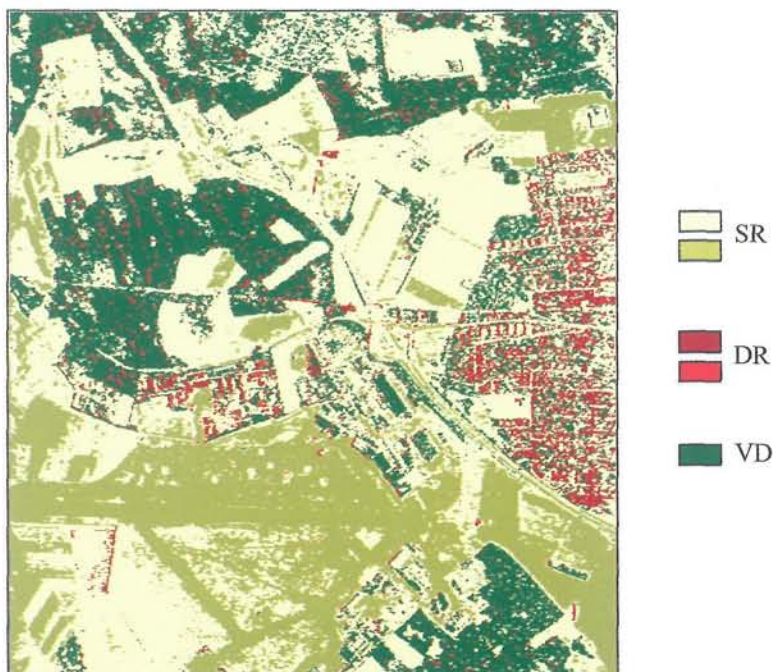


Figure 10 Selection of significant mechanisms from the H-A plane (top). Unsupervised identification to canonical scattering mechanisms (top)

The unsupervised classification results show a good discrimination of the three basic scattering mechanisms over the scene under consideration. Forested areas are well separated from the rest of the scene. Buildings, characterized by double bounce scattering, can be distinguished over the urban area and the DLR. It may be noted that the identification assigns some buildings to the volume diffusion class. The polarimetric properties as well as the power related information do not permit to separate these targets from forests. Such buildings have specific orientations with respect to the radar and particularly rough roofs and back-scatter randomly polarized waves.

#### 4. Polarimetric Interferometric SAR Data Analysis

##### 4.1. Polarimetric Interferometric Representations

Interferometric measurements of a target consist in two acquisitions of its electromagnetic response, realized from slightly different positions. The polarimetric interferometric behavior of a target is then fully described by two scattering matrices,  $S_1$  and  $S_2$ . A six element complex target vector,  $k_6$ , obtained by stacking target vectors from each interferometric image, gathers the polarimetric

interferometric information into a compact representation. The corresponding (6×6) interferometric coherency matrix is given by<sup>10-12</sup>:

$$\mathbf{T}_6 = \frac{1}{n} \sum_n \mathbf{k}_6 \mathbf{k}_6^\dagger \text{ with } \mathbf{k}_6 = \begin{bmatrix} \mathbf{k}_1 \\ \mathbf{k}_2 \end{bmatrix} \text{ and } \mathbf{k}_i = [k_{i1}, k_{i2}, k_{i3}]^T \quad (24)$$

This coherency matrix has the following structure:

$$\mathbf{T}_6 = \begin{bmatrix} \mathbf{T}_1 & \mathbf{T}_{12} \\ \mathbf{T}_{12}^\dagger & \mathbf{T}_2 \end{bmatrix} \text{ with } \mathbf{T}_{12} = \frac{1}{n} \sum_n \mathbf{k}_1 \mathbf{k}_2^\dagger \quad (25)$$

The matrices  $\mathbf{T}_1$  and  $\mathbf{T}_2$  are the standard n-look (3×3) hermitian sample covariance matrices for separate images.  $\mathbf{T}_{12}$  is a (3×3) complex matrix which contains information about the interferometric correlation between  $\mathbf{k}_1$  and  $\mathbf{k}_2$ . The target vector  $\mathbf{k}_6$  follows a complex normal zero mean distribution  $N_c(\mathbf{0}, \mathbf{\Sigma}_6)$ , with  $\mathbf{\Sigma}_6$  its (6×6) covariance matrix. The sample (6×6) covariance matrix  $\mathbf{T}_6$  has then a complex Wishart distribution  $W_c(n, \mathbf{\Sigma}_6)$ , characterized by n degrees of freedom<sup>6,16</sup>.

#### 4.2. Polarimetric Interferometric Coherence Set

The sample interferometric correlation matrix,  $\mathbf{T}_{12}$ , has complex diagonal elements, from which is computed a three polarimetric complex coherence set as follows:

$$(\gamma_1, \gamma_2, \gamma_3) \quad \text{with} \quad \gamma_i = \frac{\langle k_{1i} k_{2i}^* \rangle}{\sqrt{\langle k_{1i} k_{1i}^* \rangle \langle k_{2i} k_{2i}^* \rangle}} \quad (26)$$

where the operator  $\langle \rangle$  stands for the sum over n samples represented in (25). Standard real coherence values are obtained from the modulus of  $\gamma_i$ , while its arguments correspond to the interferometric phase difference. It may be noted that the coherence defined in (26) is not invariant under a change of polarimetric basis. In general, coherence may be decomposed into multiplicative contributions as<sup>10</sup>:

$$\gamma = \gamma_{SNR} \cdot \gamma_{spatial} \cdot \gamma_{temporal} \cdot \gamma_{polar} \quad (27)$$

where the different terms indicate decorrelations related respectively to the back-scattered wave signal to noise ration, the spatial distribution of the illuminated scatterers, temporal variations between the acquisitions and finally the polarization state. Coherence images for the different polarization channels are displayed in Figure 9.

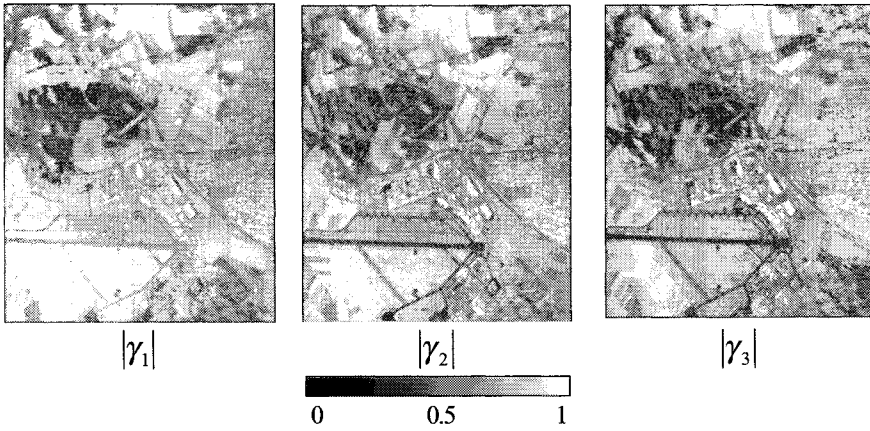


Figure 11 Interferometric coherences for the different polarimetric channels in the Pauli basis

Range filtering and topographic phase removal procedures are applied to the interferometric data sets prior to the computation of the polarimetric interferometric coherences. The range filtering procedure corrects wave number shifts inherent to interferometric measurements.

High coherence values are observed for surface areas in the co-polarized channels. Smooth surfaces, like the runway, scatter waves in the radar direction with a low signal to noise ratio, involving a low coherence in every polarization channel. In a general way, the maximum of the three coherence set corresponds to the first or second channel according to the dominant scattering mechanism, i.e. single or double bounce reflection respectively. Over forested areas, all three coherencies have low values.

Cloude and Papathanassiou<sup>10-12</sup> introduced the following original formulation of the polarimetric interferometric coherence:

$$\gamma_i = \frac{\mathbf{w}_i^T \mathbf{T}_{12} \mathbf{w}_2}{\sqrt{\mathbf{w}_1^T \mathbf{T}_1 \mathbf{w}_1 \quad \mathbf{w}_2^T \mathbf{T}_2 \mathbf{w}_2}} \quad (28)$$

where  $\mathbf{w}_1$  and  $\mathbf{w}_2$  are three elements complex vectors. The formulation of (27) permits to compute the interferometric coherence of a polarization channel in any emitting-receiving polarization basis.

Papathanassiou and Cloude further developed this concept to define an optimal coherence set  $(\gamma_{opt1}, \gamma_{opt2}, \gamma_{opt3})$ , with  $\gamma_{opt1} \geq \gamma_{opt2} \geq \gamma_{opt3}$ . The optimal coherence set is obtained by analytically tuning the projection vectors,  $\mathbf{w}_1$  and  $\mathbf{w}_2$ , to maximize the modulus of the coherences in 3-dimensional, 2-dimensional and 1-dimensional spaces.

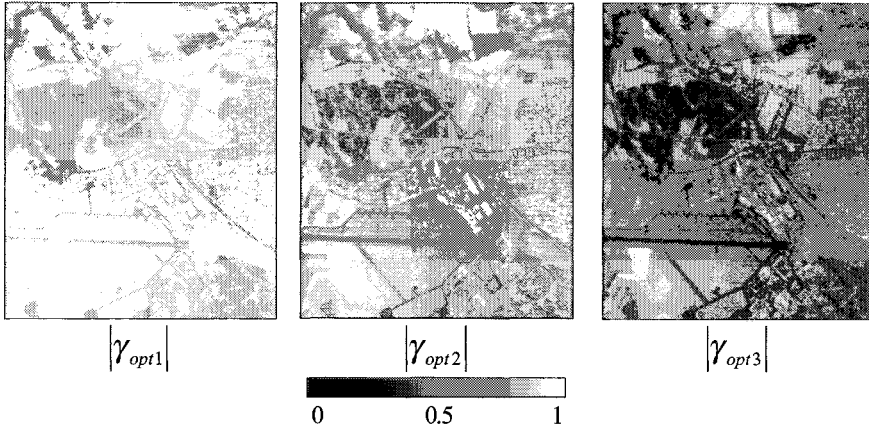


Figure 12 Optimal interferometric coherences

The results of the optimization procedure presented in Figure 12 show an enhanced contrast between the different optimal coherences. The first one has values close to one over the major part of the considered scene and intermediate values over forested areas and low SNR targets. The third one shows minimal values over decorrelating media such as forest and smooth surfaces and reaches high values for a limited amount of very coherent point scatterers. The complete optimized coherence set represents highly descriptive indicators of the polarimetric interferometric properties of each natural media and may then be used efficiently in a classification process.

## 5. Unsupervised Classification of Polarimetric Interferometric SAR Data

As has been mentioned, the optimized coherency set offers a high degree of description of the coherent properties of a medium with respect to the polarization. The color-coded image presented in Figure 13 represents the joint information associated to the optimal coherencies.





Figure 13 Color-coded image of the optimal coherence set

The color-coding used for the joint representation of the optimal coherency set reveals particular behaviors of different types of natural medium under examination.

White areas indicate targets showing high coherence independently of the polarization. Such a behavior is characteristic of point scatterers and bare soils. They also correspond, to buildings, fences and some vegetation free agricultural fields.

Green zones reveal the presence of a single dominant coherent mechanism within the resolution cell. Secondary coherences, associated to the red and blue channels have significantly lower values. Such zones correspond to surfaces with low SNR responses and some particular fields.

Forested areas, characterized by a dark green color have scattering features dominated by a single mechanism but with a very low coherence.

A comparison of the image of Figure 13 with the polarimetric color-coded image shown in Figure 1 indicates that the distribution of strictly polarimetric and polarimetric interferometric features over surfaces and agricultural fields are significantly different. Coherence related information permits to discriminate particular buildings that cannot be separated from forested areas using only polarimetric data.

Over forested areas, the polarimetric color-coded image shows homogeneous zones, while interferometric data indicate that there exist large variations of the coherent scattering properties corresponding to clear-cuts and low-density forest. In order to isolate the polarization dependant part of the optimal coherencies, it is necessary to define their relative values as:

$$\tilde{\gamma}_{opt i} = |\gamma_{opt i}| / \sum |\gamma_{opt j}| \quad \text{with} \quad \tilde{\gamma}_{opt 1} \geq \tilde{\gamma}_{opt 2} \geq \tilde{\gamma}_{opt 3} \quad (29)$$

The relative optimal coherence spectrum can be fully described by two parameters. We propose to define  $A_1$  and  $A_2$  as characteristic indicators of the distribution of the coherency in the different optimized channels.

$$A_1 = \frac{\tilde{\gamma}_{opt 1} - \tilde{\gamma}_{opt 2}}{\tilde{\gamma}_{opt 1}} \quad \text{and} \quad A_2 = \frac{\tilde{\gamma}_{opt 1} - \tilde{\gamma}_{opt 3}}{\tilde{\gamma}_{opt 1}} \quad (30)$$

These parameters indicate relative amplitude variations between the different optimized channels. Similarly to the polarimetric case, the indicators  $A_1$  and  $A_2$  may be used to estimate the number of independent coherent scattering mechanisms from the optimization results. The different optimal coherence set configurations are represented and identified in Figure 14.

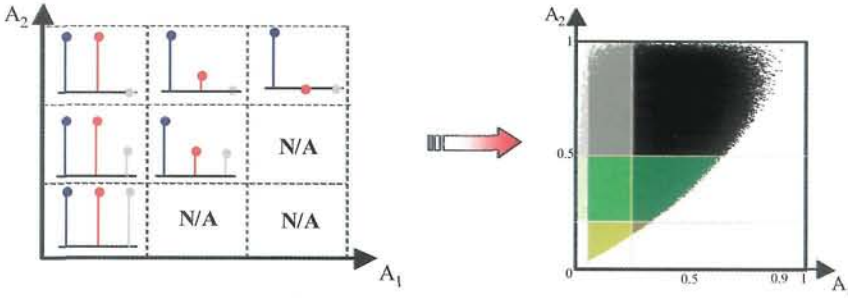


Figure 14 Discrimination of different optimal coherence set using  $A_1$  and  $A_2$  (left). Selection in the  $A_1$ - $A_2$  plane (right)

The schematic on the left hand side of Figure 14 separates the different optimal coherence set configurations in five classes. The diagonal classes correspond to configurations for which  $\tilde{\gamma}_{opt 2} = \tilde{\gamma}_{opt 3}$  with different importance with respect to the largest normalized coherence  $\tilde{\gamma}_{opt 1}$ . A column describes configurations with a constant proportionality ratio between  $\tilde{\gamma}_{opt 2}$  and  $\tilde{\gamma}_{opt 1}$ . The real segmentation from  $A_1$  and  $A_2$  is realized over nine classes in order to improve the resulting clusters descriptivity and accuracy. Results of the normalized optimal coherence set classification are shown in Figure 15.

Approximately four major classes arise from the identification of the optimal coherence set distribution. This unsupervised segmentation was also found to

achieve a high degree of descriptivity on other scenes observed with different baselines<sup>15</sup>. This is a consequence of both the coherence optimization and the definition of a relative coherence set.

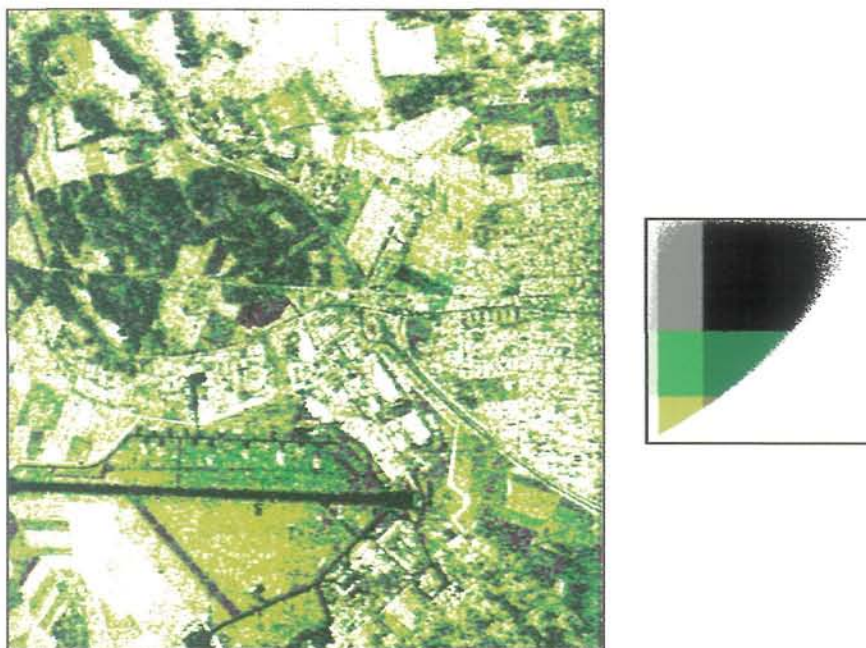


Figure 15 Unsupervised identification of the number of coherent scattering mechanisms.

The classification results are used to provide an adequate initialization to a segmentation merging polarimetric and interferometric analysis results. The classification algorithm processes the different canonical scattering mechanisms separately. Pixels belonging to one of the typical scattering type shown in Figure 10 are first segmented using the results of the interferometric coherence set identification depicted in Figure 14. The resulting clusters are used to initialize a k-mean unsupervised segmentation procedure based on the Wishart distribution of the  $(6 \times 6)$  polarimetric interferometric coherency matrix,  $\mathbf{T}_6$ , introduced in (24). In this way, pixels are segmented according to their polarimetric and interferometric features. Results for the Volume Diffusion and Surface Reflection classes are shown in Figure 16.





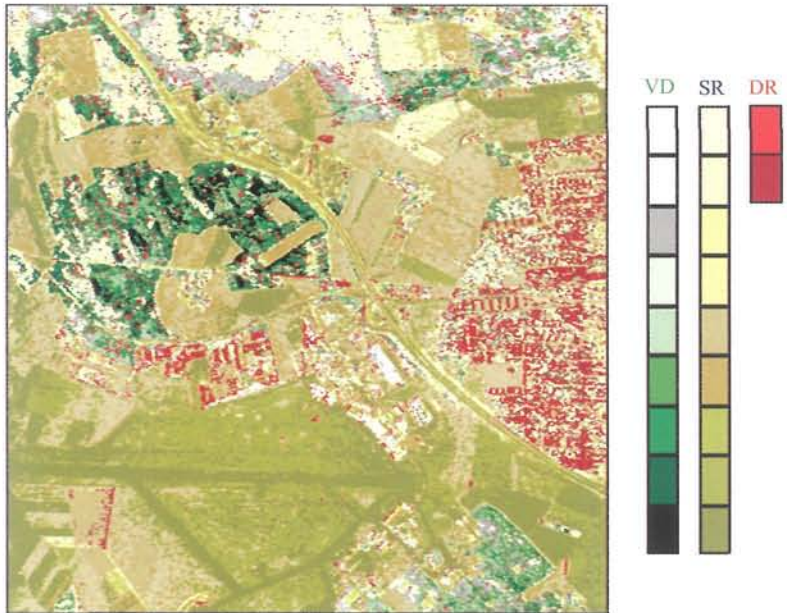


Figure 16 Unsupervised polarimetric interferometric segmentation results over the Volume Diffusion basic class (top), Surface Reflection class (middle) and all three canonical scattering mechanisms (bottom)

Clusters resulting from the ML segmentation are assigned a color indicating their average coherence, ranging from black for low coherence to white for high coherence.

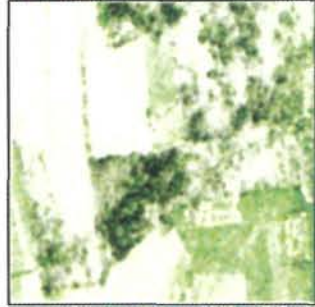
Globally, polarimetric interferometric characteristics are efficiently segmented into compact clusters corresponding to scatterers with similar polarimetric and interferometric characteristics.

The segmentation of the Volume Diffusion class successfully discriminates buildings, dense forest, sparse forest and clear-cuts. Surface Reflection areas are separated into segments according to both polarimetric and interferometric characteristics information. Details of the classification are displayed in Figure 17 for two particular zones corresponding to the DLR buildings and forest parcels.

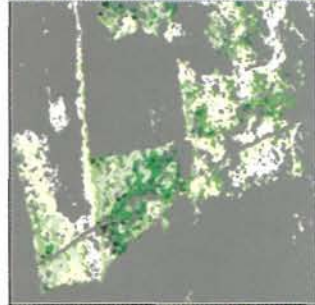
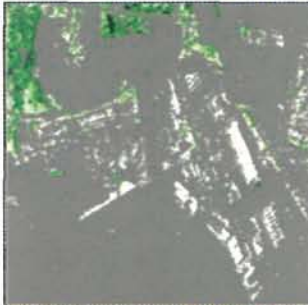
a)



b)



c)



d)





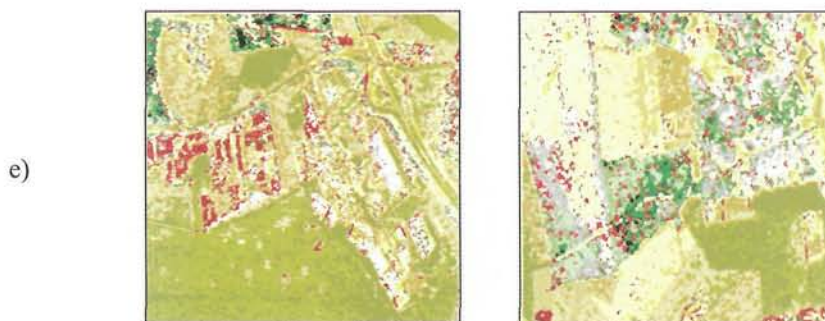


Figure 17 Polarimetric Interferometric classification results over two areas. a) polarimetric color coded image b) optimal coherence set color coded image c) volume diffusion classification results d) surface reflection classification results e) Global classification into three canonical mechanisms including double bounce reflection

It may be seen in Figure 17 a) that the DLR buildings do not have the same polarimetric behavior. Some of them have a polarimetric response dominated by double bounce scattering while the rest present behaviors characteristic of volume diffusion with high back-scattered power and significant amplitude in the cross-polarized channel. The color-coded interferometric coherence image in Figure 17 b) indicates that scattering mechanisms over the buildings have a high degree of coherence. The classification approach proposed in this paper first assigns the different buildings to the double bounce reflection or volume diffusion classes according to their polarimetric behavior. The introduction of the interferometric information permits then to separate vegetated areas from man-made targets and corrects possible errors in the interpretation of the scattering from strictly polarimetric data only, Figure 17 c). The classification of the areas dominated by surface reflection, presented in Figure 17 d) clearly discriminates very smooth surfaces with low SNR, car parks and agricultural fields.

The selected forested area presents a uniform polarimetric behavior dominated by volume diffusion characteristic of dense vegetation observed at L band. The color-coded interferometric image shows some large variations of the polarimetric interferometric coherence set. In some parts of the studied area, optimal emitting and receiving polarization states lead to a high degree of coherence. The joint use of polarimetric and interferometric data permits to segment the corresponding sparsely vegetated zones or clear-cuts and provides a final mapping with significantly increased information content.

## 6. Conclusion

This chapter presents an unsupervised procedure for classifying polarimetric interferometric SAR data. Data acquired in polarimetric and interferometric modes have complementary characteristics; their joint use in a classification process

provides significantly higher performance. The classification approach resides on separate polarimetric and interferometric data classification and interpretation.

Each scatterer is accurately identified to a basic scattering mechanism using efficient polarimetric indicators. The results of this identification are then used to evaluate polarimetric properties of clusters obtained from an unsupervised statistical segmentation based on the multivariate Wishart distribution.

A parameterization of an optimal interferometric coherence spectrum is used to segment data according to their interferometric properties.

Finally, an unsupervised classification process, gathering polarimetric and interferometric results, is applied to each canonical scattering mechanism. The resulting images show significant improvements compared to the strictly polarimetric case. Clear-cuts sparse and dense forests are separated according to their coherent properties. Particular buildings having a polarimetric behavior similar to forest are discriminated. Surfaces are also better classified with more descriptive and more informative clusters.

It is important to note the generality of this classification method. The parameters used in these studies were chosen so as to reduce the sensitivity of the whole algorithm to changes from one site to the other or for different measurement conditions. Its applications to data acquired over a different site with a different baseline also led to equally satisfying results.

## 7. References

1. E. Rignot, R. Chellappa and P. Dubois, "Unsupervised Segmentation of Polarimetric SAR Data Using the Covariance Matrix" *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 30, no. 4, pp. 697-705, July 1992.
2. J.J. van Zyl, C.F. Burnette "Bayesian classification of polarimetric SAR images using adaptive a-priori probabilities", *International Journal of Remote Sensing*, vol. 13, pp. 835-840, 1992.
3. J.A. Kong, S. H. Yueh, R.T. Shin, J.J. van Zyl, "Classification of Earth Terrain using Polarimetric Synthetic Aperture Radar Images", Chapter 6 in *PIER Volume 3*, ed. J.A. Kong Elsevier 1990
4. J.S. Lee, M.R. Grunes, R. Kwok, "Classification of multi-look polarimetric SAR imagery based on the complex Wishart distribution", *International Journal of Remote Sensing*, vol. 15, No. 11, pp. 2299-2311, 1994.
5. K.S. Chen, W.P Huang, D. H. Tsay, F. Amar, "Classification of Multifrequency Polarimetric SAR Image Using a Dynamic Learning Neural Network," *IEEE Trans. Geoscience and Remote Sensing*, vol. 34, no.3, 814-820, 1996.
6. L. Ferro-Famil, E. Pottier, J. S. Lee, "Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/Alpha-Wishart classifier", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, n°11, pp. 2332-2342, November 2001.



7. S. R. Cloude, E. Pottier, "A Review of Target Decomposition Theorems in Radar Polarimetry", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, n°2, pp. 498-518, September 1995.
8. S. R. Cloude, E. Pottier, "An Entropy Based Classification Scheme for Land Applications of Polarimetric SAR", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, n°1, pp. 68-78, January 1997.
9. A Freeman, S. Durden, "A three component scattering model for polarimetric SAR data", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, n°3, pp. 963-973, May 1998.
10. Papathanassiou, K. P., *Polarimetric SAR Interferometry*, 1999, Ph. D. Thesis, Tech. Univ. Graz (ISSN 1434-8485 ISRN DLR-FB-99-07).
11. Cloude, S. and K. Papathanassiou, 1998, "Polarimetric SAR Interferometry", *IEEE Trans. Geosci. Remote Sensing*, vol. 36 (4), pp. 1551-1565.
12. Papathanassiou, K. P. and S. Cloude, 2001, "Single-baseline polarimetric SAR interferometry", *IEEE Trans. GRS*. 39 (6), 2352-2363, November 2001
13. J.S. Lee, M.R. Grunes, T.L. Ainsworth, L. Du, D.L. Schuler and S.R. Cloude, "Unsupervised Classification of Polarimetric SAR Images by Applying Target Decomposition and Complex Wishart Distribution", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2249-2258, Sept. 1999, also *Proceedings of the Fourth International Workshop on Radar Polarimetry*, PIERS 1998, Nantes, France, 13-17 July 1998.
14. E. Pottier, J.S. Lee, "Unsupervised Classification of PolSAR images based on the Complex Wishart Distribution and the H/A/ $\alpha$  Polarimetric Decomposition Theorem", *EUSAR*, Munich, Germany, May 2000.
15. L. Ferro-Famil, E. Pottier, J.S. Lee, "Classification and Interpretation of Polarimetric SAR data", *Proceedings of IGARSS*, June 2002, Toronto, Canada.
16. L. Ferro-Famil, 2000, "*Multi-frequency and multi-temporal remote sensing of natural media using polarimetric SAR data*", Ph. D. Thesis, IRESTE, University of Nantes, France, December 2000.
17. W.M. Boerner et al., "Polarimetry in Radar Remote Sensing: Basic and Applied Concepts" Chapter 5 *Principles and Applications of Imaging Radar, The Manual of Remote Sensing*, 3rd Edition, The American Society for Photogrammetry and Remote Sensing, March 1998.
18. J.R. Huynen, "*Phenomenological theory of radar targets*", Ph. D. dissertation, Drukkerij Bronder-offset, N.V. Rotterdam, 1970.
19. E. Pottier, W.M. Boerner, D.L. Schuler, "Polarimetric Detection and Estimation of Ship Wakes", *Proceedings of IGARSS*, July 1999, Hamburg, Germany.
20. N. R. Goodman, "Statistical analysis based on a certain multi-variate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, pp. 152-177, 1963.

This page is intentionally left blank

# **Wavelet Analysis and Applications**

This page is intentionally left blank

## CHAPTER 7

# WAVELET ANALYSIS OF SATELLITE IMAGES IN OCEAN APPLICATIONS

Antony K. Liu, Sunny Y. Wu, and Yunhe Zhao

*Oceans and Ice Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771*  
*E-mail: [liu@neptune.gsfc.nasa.gov](mailto:liu@neptune.gsfc.nasa.gov)*

A two-dimensional wavelet transform is a very efficient bandpass filter, which can be used to separate various scales of processes and show their relative phase/location. A feature tracking procedure based on wavelet transform has been developed and used for image processing at NASA Goddard Space Flight Center for the past several years. The two-dimensional Gaussian wavelet has been applied to satellite images for coastal monitoring (e.g. oil spills) and for ice edge and ice floe tracking from synthetic aperture radar (SAR), ocean color, and infrared (IR) data. However, SAR is valuable for feature tracking due to the fine spatial resolution of the data, but its less than daily coverage may be a serious problem for some ocean applications. A similar technique of wavelet analysis for scatterometer and radiometer data has been developed to obtain daily sea ice drift information in the Arctic region. This technique provides improved spatial coverage and better temporal resolution over techniques utilizing data from SAR. From low earth orbits, ocean surface feature tracking analyses have always been based on data from a single orbital sensor collected over the revisit interval of a single satellite. For the first time, ocean surface layer currents have been derived by wavelet feature tracking of ocean color data from different sensors on different satellites. Ocean color data can be used as a tracer for measuring ocean surface layer currents, because the ocean color signal comprises information from a deeper water depth than surface signatures. The results of feature tracking from these multiple sensors demonstrate that wavelet analysis of satellite data is a very useful tool for image processing.

### 1. Introduction

Basically, wavelet transforms are analogous to Fourier transform but are localized both in frequency and time. A two-dimensional wavelet transform is a highly efficient band-pass filter, which can be used to separate various scales of processes and show their relative phase/location [1]. A feature tracking procedure based on wavelet transform has been developed and used for image processing at NASA Goddard Space Flight Center for the past several years. The two-dimensional Gaussian wavelet (often referred to as a "Mexican hat" wavelet) has been applied to satellite images to separate various scale processes including relative phase/location information for coastal monitoring [2] and for ice edge and ice floe tracking [3]. This Gaussian wavelet transform works as a band-passed filter. Ideally, only features with length scale within a chosen band are retained upon transformation [4]. The wavelet transforms of SAR images can be used for near real-time "quick look" screening of satellite data (feature detection), data reduction (binary image), and image enhancement (edge linking).

However, SAR is valuable for feature tracking due to the fine spatial resolution of the data, but its less than daily coverage may be a serious problem for some ocean applications. A similar procedure of wavelet analysis for scatterometer and radiometer data can be used to obtain daily sea ice drift information for the Arctic region [5],[6]. This technique provides improved spatial coverage over the existing array of Arctic Ocean buoys and better temporal resolution over techniques utilizing data from SAR. Comparisons with ice motion derived from ocean buoys give good quantitative agreement. Both comparison results from scatterometer and radiometer are compatible, and both results are definitely complement with each other. Then these sea-ice drift daily results from satellite data can be merged as a composite map by some data fusion techniques. Examples of derived ice-drift maps illustrate large-scale circulation reversals over a period of few days.

Historically, from low earth orbits, ocean surface feature tracking analyses have been based on data from a single orbital sensor collected over the revisit interval of a single satellite. For the first time, ocean surface layer currents have been derived by wavelet feature tracking of ocean color data from different sensors on different satellites [7]. Satellite ocean color data provide important insight into the marine biosphere by quantifying certain fundamental properties on a global scale. In addition, satellite ocean color data can also be used as a tracer for measuring ocean surface layer currents by wavelet feature tracking, because the ocean color signal comprises information from a deeper water depth than surface signatures (such as sea surface temperature).

In this chapter, some recent works developed at NASA Goddard Space Flight Center using wavelet analysis for satellite image processing are reviewed. After a general introduction section, the contents are organized as follows. An introduction to the satellites and sensors used later in the case studies is first presented. A brief description of the two-dimensional wavelet analysis, which forms the basis of our tracking technique, is presented next. This is followed by several case studies of feature extraction based primarily on the wide swath SAR data to demonstrate the many applications such as monitoring of fronts, sea ice boundary, and oil spills. Case studies employing feature tracking technique from multiple sensors where sea-ice drift is derived from sequential SAR data, sea-ice motion from QuikSCAT and SSM/I data, and ocean surface layer drift from MODIS and SeaWiFS data are then presented. The concluding remarks are provided in the last section.

## **2. Recent Satellites and Sensors for Ocean Applications**

In this section, some recent satellites and sensors that used for ocean applications are introduced as background for the case studies presented later in this chapter.

### **a) Synthetic Aperture Radar (SAR):**

Synthetic aperture radar (SAR) is a side-looking imaging radar usually operating on either an aircraft or a spacecraft. SAR instruments transmit radar signal, thus

providing their own illumination, and then measure the strength and phase of the signals scattered back to the instrument. Radar waves have much longer wavelengths compared with light, allowing them to penetrate clouds with little distortion. The ability of SAR for monitoring surface signatures of swells, wind fronts, and eddies has been amply demonstrated. With all-weather, day/night imaging capability, SAR can penetrate clouds, smoke, haze and darkness to acquire high quality images of the Earth's surface. This frequently makes SAR the sensor of choice for many overcast coastal regions. Space agencies of the U.S., Canada, and Europe use SAR imagery on an operational basis for sea ice monitoring, and for the detection of icebergs, ships, and oil spills. However, there can be considerable ambiguity in the interpretation of physical processes responsible for the observed ocean features [8].

The first spaceborne SAR was flown on the U.S. satellite Seasat in 1978. Although Seasat only lasted three months, analysis of its data confirmed the sensitivity of SAR to the geometry of surface features. It was followed by a decade with very little progress in SAR technology. On March 31, 1991 the Soviet Union became the next country to operate an earth-orbiting SAR with the launch of Almaz-1. Almaz-1 returned to earth in 1992 after operating for about 18 months. The European Space Agency (ESA) launched its first remote sensing satellite, ERS-1, with a C-band SAR on July 17, 1991. Shortly thereafter, the JERS-1 satellite, developed by the National Space Development Agency of Japan (NASDA), was launched on February 11, 1992 with an L-band SAR. This was followed a few years later by ERS-2, launched in April 1995 by ESA, and RADARSAT-1, the first Canadian remote sensing satellite, launched in November 1995 by the Canadian Space Agency (CSA). RADARSAT-1 has a ScanSAR mode with a 500 km swath and a 100 m resolution, an innovative variation of the conventional SAR (with a swath of 100 km and a resolution of 25 m). Most recently, ESA's Envisat, equipped with an Advanced SAR (ASAR), was launch on March 1, 2002. In the near future, CSA has RADARSAT-2 scheduled for launch in early 2004, and NASDA has the Advanced Land Observing Satellite (ALOS), which was equipped with a Phased Array type L-band SAR (PALSAR), scheduled for launch in the summer of 2004.

b) QuikSCAT:

QuikSCAT, a "quick recovery" mission to fill in the gap created by the loss of data from NSCAT, when the ADEOS-1 satellite lost power in June, 1997, was launched on June 19, 1999. It will continue to add to the important ocean wind data set begun by NSCAT in September, 1996. QuikSCAT has 25km resolution with 5km resolution from enhanced processing. The finer resolution will further reduce the uncertainty of the ice motion product that inherits from satellite data and therefore may improve the accuracy of the product.

c) Special Sensor Microwave/Imager (SSM/I):

The SSM/I is a seven-channel, four frequency, linear-polarized, passive microwave radiometric system which measures atmospheric, ocean, ice and terrain microwave brightness temperature at 19, 22, 37 and 85 GHz. The 85GHz data are used in this study because of its 12.5 km resolution. The Advanced Microwave Radiometer for

EOS (AMSR-E) was put to use aboard the satellite Aqua upon its launch on May 4, 2002. Spatial resolution of the individual measurements is 5.4 km at 89 GHz.

d) **Sea-viewing Wide Field-of-view Sensor (SeaWiFS):**

The spacecraft SeaStar carries the SeaWiFS instrument and was launched to low earth orbit on board an extended Pegasus launch vehicle on August 1, 1997. The purpose of SeaWiFS Project is to provide quantitative data on global ocean bio-optical properties to the Earth science community. Subtle changes in ocean color signify various types and quantities of marine phytoplankton (microscopic marine plants), the knowledge of which has both scientific and practical applications.

e) **Moderate Resolution Imaging Spectroradiometer (MODIS):**

MODIS is a 36-band spectroradiometer measuring visible and infrared radiation and obtaining data that will be used to derive products ranging from vegetation, land surface cover, and ocean chlorophyll fluorescence to cloud and aerosol properties, fire occurrence, snow cover on the land, and sea ice cover on the oceans. The first MODIS instrument was launched on board the Terra satellite in December 1999 and the second MODIS instrument was launched on board the Aqua satellite in May 2002.

### **3. Wavelet Analysis**

#### **3.1 A Histogram Screening Routine for Feature Detection**

Intuitively, if one is to discern a “feature” from an image, there must be sufficient contrasts in the pixel grayscales. Statistically, that means the pixels that made up the scene do not all belong to a single group, but rather, to multiple distinct groups, which translates into multiple peaks in histogram presentation. The histogram screening routine is based on this physical intuition [4]. It further assumes that multiple peaks in histogram is, in addition to being a necessary condition, also a sufficient condition for the existence of a visually distinguishable feature(s).

In practice, the histogram screening routine first divides an image into smaller, overlapping regions. Each region shares 50% of the same pixels with each of its four neighboring regions; consequently each pixel may belong to four different regions at most. The intensity histogram is calculated for each region. Overlapping of the screening area is necessary to ensure a feature, such as a front, that happens to follow the boundary of two adjacent screening regions will not be overlooked. If more than one distinct peaks are detected, the region under screening is deemed to be one with feature(s), otherwise to be one without a feature. A matrix corresponding to the original image is constructed with each grid point representing an individual screening area and its value the screening result (“1” for feature found; “0” for feature not found).



In examining the histogram screening result, an eight-neighbor connection scheme is adapted, that is, a grid node is considered to be connected to its eight immediate neighbors: its upper left, upper, upper right, left, right, lower left, lower and lower right grid nodes. In this way, the linear dimension and the area of a feature can be estimated, thus to decide whether the feature is large enough for a particular study to render further investigation. Once a feature of significant size is detected, one can use a feature extraction routine, such as the two-dimensional wavelet analysis described in the following section, to delineate it.

### 3.2 Wavelet Transform

In general, two-dimensional continuous wavelet transform  $W_s(a, \underline{b})$  of a function  $s(\underline{r})$ , where  $\underline{r} = (x, y)$ , is expressed in terms of the complex valued wavelet function  $w(\underline{r})$  as follows:

$$W_s(a, \underline{b}) = \frac{1}{a} \int s(\underline{r}) w^* \left( \frac{\underline{r} - \underline{b}}{a} \right) d\underline{r} \quad (1)$$

in which the wavelet function is dilated by the factor  $a$ , and shifted by the vector  $\underline{b}$ . The function  $w(\underline{r})$  is the basic wavelet [1]. The superscript \* indicates complex conjugate. Two frequently used wavelet functions are the Morlet wavelet, a Gaussian modulated sine and cosine wave packet, and the Mexican hat, the second derivative of a Gaussian function. In this study, we use the Mexican hat that can be expressed as follows:

$$w(x, y) = (2 - (x^2 + y^2)) \exp\left(-\frac{x^2 + y^2}{2}\right). \quad (2)$$

Since convolution is commutative with respect to differentiation, the resulting wavelet transform is the Laplacian of a Gaussian smoothed function. Thus, zeroes correspond to the inflection points of the original function. The contours of zero crossing indicate the edges in the pattern of the input function.

To perform the wavelet transformation, first a suitable  $a$  value is chosen, which corresponds to the length scale of the Gaussian function. The general rule of thumb is to choose this value based on 1) the length scale of the feature of interest; and 2) the resolution (or pixel spacing) of the original image.

### 3.3 Threshold for Feature Extraction

A differential detection is then carried out to determine the pixel locations of significant differentials so that the feature of interest can be delineated from the background. In this chapter, both kinds of wavelet-transformed data ( $w1$ , the first derivative and  $w2$ , the second derivative) are examined. An edge is expected to locate at pixel  $(x_0, y_0)$  if both of the following conditions are met:

$w1(x_0, y_0)$  exceeds a chosen threshold value  $T$ ,

where  $T = [\overline{w1(x, y)}] + t \times [w1(x, y)']$ , and

$[\overline{w}]$ : the ensemble average;

$[w]$  : the standard deviation of the ensemble.

$w^2(x_0, y_0)$  equals to a predetermined value  $c$  which is typically in the neighborhood of zero.

The choices of  $t$  and  $c$  rely to a great extent on the observation as to where the actual edges lie, which can be achieved by visual inspection. For noise-free images, the threshold  $t$  can be chosen such that all amplitude changes above a minimum contrast level are detected as edges, and all others are considered non-edges. With noisy images, such as SAR with speckle noise, the threshold selection becomes a tradeoff between missing valid edges and designating noise-induced false edges.

The Laplacian  $G(x,y)$  of the original image  $F(x,y)$  is zero if  $F(x,y)$  is constant or changing linearly in amplitude. If the rate of change of  $F(x,y)$  is greater than linear,  $G(x,y)$  exhibits a sign change at the point of inflection of  $F(x,y)$ . The zero crossing of  $G(x,y)$  thus indicates the presence of an edge. In reality, since SAR imagery is subject to heavy "speckle noises", the rate of the change of the original image ( $F(x,y)$ ) equals to zero at many pixel locations throughout the image. The zero crossings of  $G(x,y)$  under this circumstance are clearly false edges. Therefore, it often becomes necessary to find a suitable value other than zero for  $c$ . By properly selecting a non-zero contour value, Liu *et. al.* (1997a) has shown that one can pick out the valid edges while reducing the amount of the false edges by as much as 80%. Obviously, in order to keep the edges identified in a sufficiently narrow band of the point of the inflection of  $F(x,y)$ , the contour value has to be sufficiently close to zero.

#### 4. Feature Extraction: SAR for Coastal Monitoring

The wavelet analysis of SAR images can be used for near real-time "quick look" screening of satellite data (feature detection), data reduction (binary image), and image enhancement (edge linking). Figure 1 shows the flow chart of feature extraction using wavelet analysis based on SAR, IR, and ocean color data as reference.

##### 4.1 An Oil Spill off Point Barrow, Alaska

The reason why oil spills are detectable on radar images is that surface films have a dampening effect on short surface waves. Radar is remarkably sensitive to small changes in the roughness of the sea. The dark appearance on radar images is due to the smooth ocean surface for very low intensity of backscattering. It is the distinctive shape and sharp boundary of surface film which enables them to be identified with a high degree of certainty.

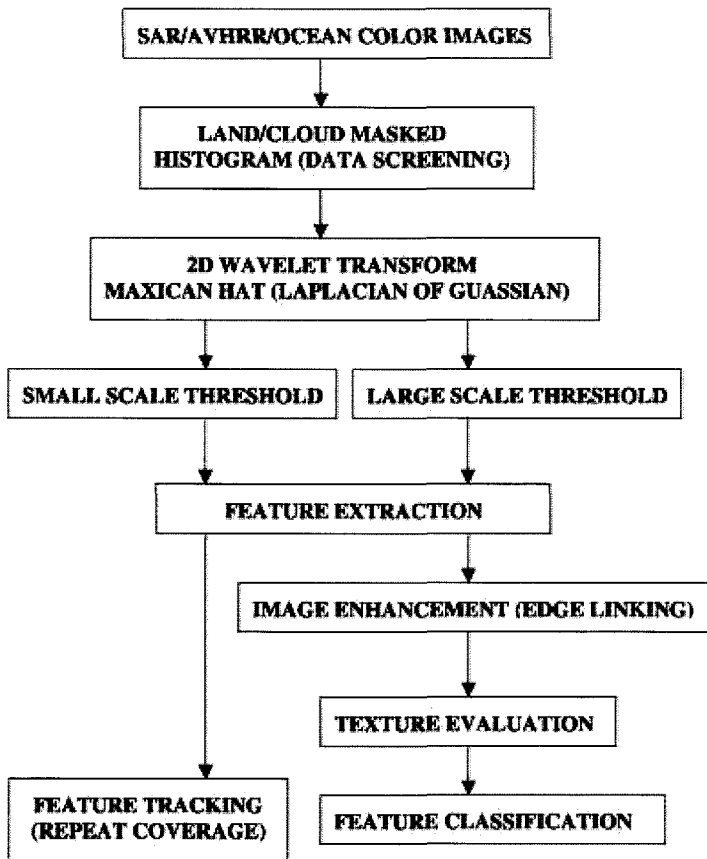


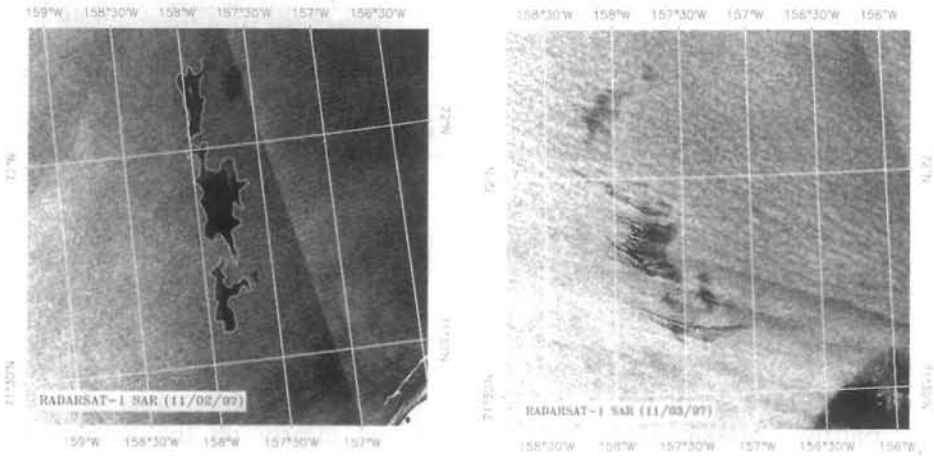
Figure 1 Flow Chart of Feature Extraction Technique Using Wavelet Analysis of Satellite Data from SAR, Ocean Color, and IR

In early November 1997, shortly after the satellite RADARSAT had resumed its coverage of the Arctic region, the SAR sensor on board captured an oil spill off Point Barrow, Alaska. The oil slicks show up clearly on the RADARSAT ScanSAR imagery on November 2, 3 and 9. The oil spill might be related to the Alaskan Oil Pipeline, although the exact source remains unidentified. Three subscenes of  $1024 \times 1024$  pixels with 100 m resolution encompassing the oil slicks was cropped out from each of the original ScanSAR images for analysis (Figure 2a, b, c).

Among the three SAR images, the oil slicks on November 2 have the most well defined shape, apparently as the result of the low wind and calm sea conditions. The matching  $a$  value for the wavelet analysis is 8, corresponding to a length scale of  $L = 800$  m. The threshold  $t$  is chosen to be 0.6. Figure 2a shows

the result of wavelet analysis with the oil slicks delineated. The oil spill shows up as three separate patches that lined up in the north-south direction.

On November 3, the wind had obviously become stronger with wind streaks clearly visible as shown in Figure 2b. The wind streaks from the SAR image follow the northwest to southeast direction. The oil slicks had broken into a series of elongated patches. A much smaller  $a = 2$  is used in order to delineate them. This value corresponds to a length scale of  $L = 200$  m. The threshold is chosen to be 2.4. In general, as  $a$  value becomes smaller, the differential detection picks up finer details. The threshold  $t$  is then raised to offset the effect. The shapes of oil slick patches are not well-defined due to the strong wind drift. Therefore, the outlines of some of the less distinguishable oil slicks are missed out as a consequence. Six days later on November 9, the oil slicks had drifted further offshore toward northwest and become two elongate patches. Figure 2c shows a subscene from the SAR image collected on November 9 with the oil slicks delineated by wavelet transform. The relevant parameters for the wavelet analysis are:  $a = 4$  for wavelet length scale, and the threshold  $t = 2.2$ . Figure 2d shows a map of Point Barrow and its vicinity to summarize the approximate locations of the oil slicks and their tracks as they were drifting further offshore in response to the wind and current.



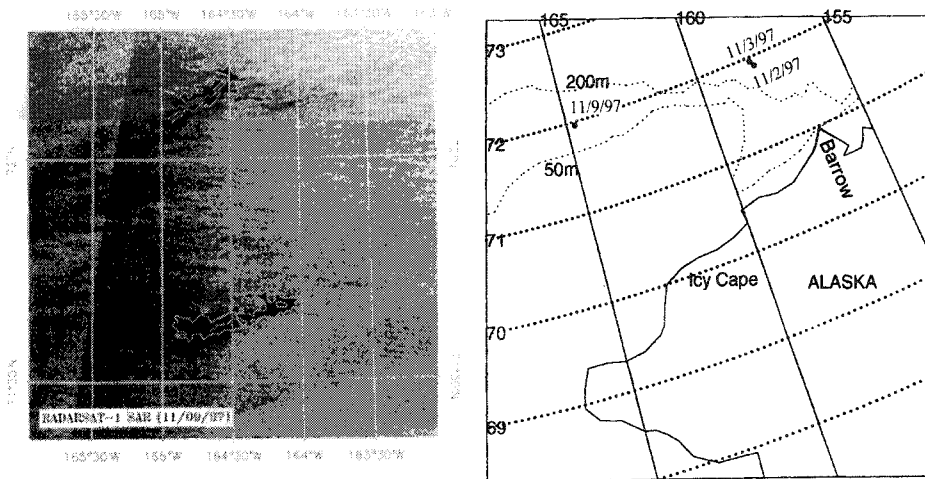


Figure 2 Oil Slicks Monitoring near Point Barrow, Alaska on (a) Nov. 2, (b) 3 and (c) 9, 1997 and (d) Location Map

## 4.2 An Ice Edge in the Bering Sea

Sea ice is not a stationary, rigid sheet of material atop of sea surface, it is driven constantly by winds and currents, and develops ridges, leads, and coastal polynyas as response to the internal stresses caused by these forcing. Not surprisingly, the marginal ice zone and the position of the ice edge are highly dynamic [9]. In the Bering Sea, for instance, the daily displacement of sea ice is easily on the order of a few kilometers, even greater during sea ice advance/retreat seasons. SAR imagery can provide valuable information on the position/motion of the ice edge.

Figure 3b is a subscene of RADARSAT ScanSAR image acquired in southeastern Bering Sea near Nunivak Island on April 3, 2000, during a period of sea ice retreat. Its aerial coverage is about 200 km by 200 km (2048×2048×100 m). Here sea ice appears heavily textured along the edges but gradually becomes smooth away from the edge. In this instance, the resulting binary image (Figure 3a) using a 128×128-pixel screening window identifies areas with heavy texture, either along the ice edge or in the interior of the sea ice where sea ice and open water are both present.

With the scaling factor  $a = 20$  (corresponding to 4km), a number of curves are identified as the most prominent boundaries at this scale by wavelet analysis. The discriminatory procedure then retains the longer ones near the border of the identified feature zone. The technique of edge linking is then applied to connect individual line segments. The white curve in Figure 3b represents the ice edge delineated by the wavelet analysis. Notice that a small open water area inside of

ice cover is also extracted by wavelet analysis due to its very low ice concentration.

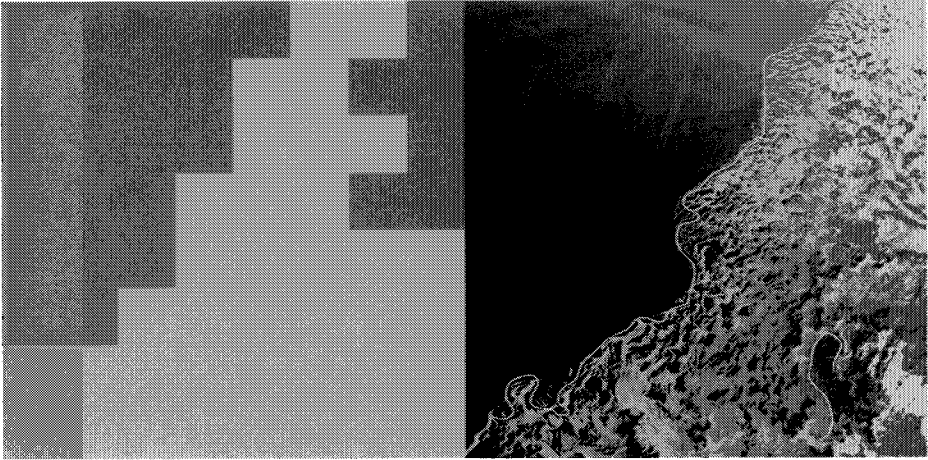


Figure 3 Marginal Ice Zone (a) Binary Image to Identify Ice Cover Area, (b) Ice Edge Extraction Using Automated Wavelet Analysis

### 4.3 Sea Ice Motion near St. Lawrence Island

St. Lawrence Island is a small island due south of Bering Strait at  $63.5^{\circ}\text{N}$ ,  $170.5^{\circ}\text{W}$ . Its extent in the east-west direction is about 150 km. Polynyas off its leeward coast, behind the tall, round-shaped Kookooligit Mountain, is a recurring phenomenon during winter. Since December of 1998, ScanSAR Wide (100-m pixel resolution, 510-km swath width) images covering St. Lawrence Island and its vicinity have been regularly received to study the evolution of the polynyas off its leeward coast as well as the surface velocity field surrounding that island. In SAR images containing floes or polynyas, the boundaries between open water and ice, or young gray ice and older brighter ice are sometimes easily identified. In such circumstances, a single continuous boundary may be detected through a single-scale wavelet transform by using the Mexican-hat wavelet. The maximum gradient change in pixel intensity along a zero-crossing contour determines the boundary location.

Two sequential images collected over St. Lawrence Island on January 4, and 6, 1999 were chosen for this study. Subscenes from the two images, each of size  $2048 \times 2048$  pixels, are shown in Figure 4a, and b, with the polynya delineated by the 2-D wavelet analysis. The shape and the make-up of the polynya have evidently changed a great deal from one scene to another over a course of 2 days, suggesting that the polynya is rapidly evolving.

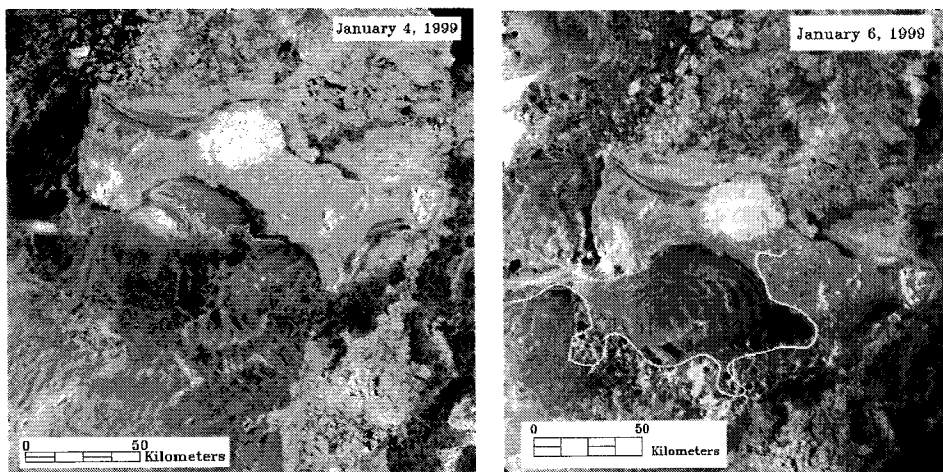


Figure 4 Two Sequential ScanSAR Images Collected over St. Lawrence Island on (a) January 4, and (b) January 6, 1999 with the Polynya Delineated by the 2-D Wavelet Analysis

Figure 5a, and b are ScanSAR subscenes of the same area just north of St. Lawrence Island on January 4, and 6. A number of ice floes in this area remain visually recognizable through out this period. A triangle, which is created by connecting the centers of mass of three particular ice floes, is indicated by a thick black line in these figures. A comparison of the two triangles on Jan. 4 (Figure 5a) and Jan. 6 (Figure 5b) indicates that the three floes had all moved slightly eastward. The displacement is by no means a simple translation. The area of the triangle had changed from approximately  $465 \text{ km}^2$  on Jan. 4 to  $275 \text{ km}^2$  on Jan. 6, indicating an area of convergence. It appears that the wind is blowing out of the west, herding the ice floes up against the more consolidated ice located on the east.

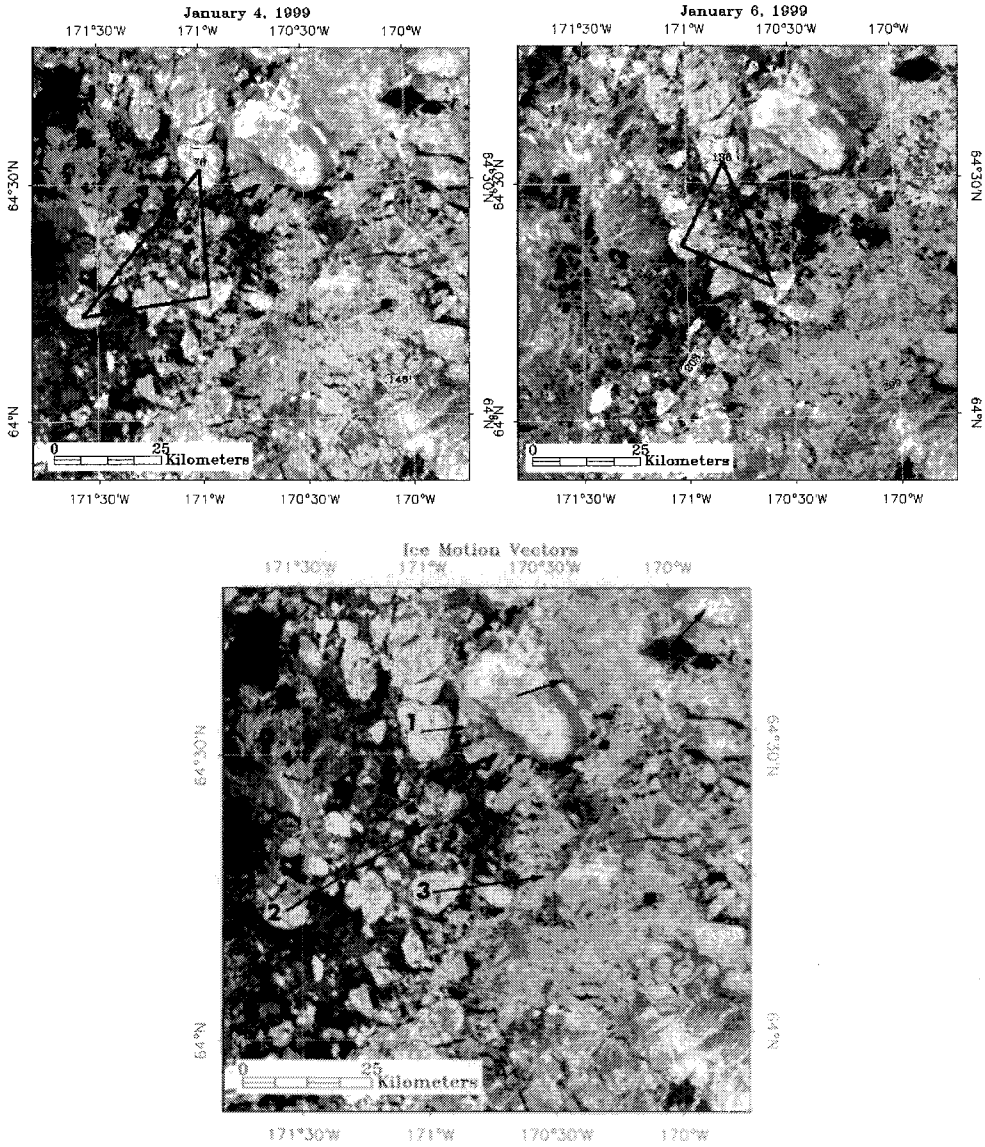


Figure 5 ScanSAR Subscene north of St. Lawrence Island on (a) January 4, and (b) January 6 1999, and (c) Ice Floe Motion during these two days.

The motion of the ice floes can be further quantified by means of 2-D wavelet analysis. First, some distinct ice floes identified by the wavelet analysis are numbered in Figure 5a, and b. By calculating the displacement of the center of mass of each floe, one can estimate the velocity of each floe between the time period that separates the two snapshots. Figure 5c shows the displacements and directions of ice floes during Jan. 4 and Jan. 6. The two floes located in the less



consolidated region have moved 10 km, or an average of 286 meters per hour, to the east. The third floe, which is located well within the consolidated region, has moved only approximated 3.4 km, or an average of 97 meters per hour, in the southeasterly direction.

Studies discussed above clearly reveal that the flow field surrounding St. Lawrence Island is fairly dynamic, both in the sense that it amounts to a considerable speed and in that the flow pattern changes quite rapidly. Results of the polynya and floe tracking suggest that the time scale associated with them is less than 5 days, presumably as a response to the synoptic weather pattern, which typically has a time scale of 3 to 5 days. Therefore, both high temporal and spatial coverage of marginal ice zone areas, such as using ScanSAR, is critical for the coastal monitoring.

## **5. Feature Tracking: Multiple Sensors**

### **5.1 Sea-Ice Motion from QuikSCAT and SSM/I**

Satellite imageries are providing the full view of sea ice motion in the polar region every day. Because of this unprecedented capability, satellite imageries have been frequently used to study polar sea ice motion and made their important contributions to the better understanding of the polar sea ice motion. Using different approaches, [5] and [6] demonstrated that sequential imagery from SSM/I 85GHz could provide ice motion observations based on wavelet transform and on cross-correlation methods, respectively. [10] reported a completely unexpected success in using NASA Scatterometer (NSCAT) to derive sea ice motion, even though NSCAT was not designed for this purpose. The combined use of multiple satellite imageries is obviously important in order to get more accurate and complete ice motion data. [11] reported that sea ice motion derived from wavelet analysis of SSM/I and NSCAT agreed quantitatively with the ice motion derived from buoy data. Thereby the ice motion derived from these two satellite data sets is compatible with and complement to that derived by buoy data. Furthermore, the two sea ice motion daily results from satellite data can be merged with those from buoys by some data fusion techniques to provide a composite map with more complete coverage of sea ice motion. Validation of sea ice motion from QuikSCAT with those from SSM/I and buoy has been studied by [12].

For reference, Figure 6 shows the flow chart of feature tracking procedure using wavelet analysis for satellite data from SSM/I, and QuikSCAT/SeaWinds. In this tracking procedure, daily satellite images of the entire Arctic region are first constructed and interpolated to fit into the numerical grid where land is masked out. Wavelet transform defined by Equations (1) and (2) is then applied to the interpolated satellite images at various scales to separate different ice textures or features. The choice of the scales for wavelet transform depends on the physical scales of the ice signatures (brightness temperature for SSM/I and backscatter/roughness for QuikSCAT) to be extracted. The effect of this wavelet transform is a band pass-filter with a threshold for feature detection. Then two

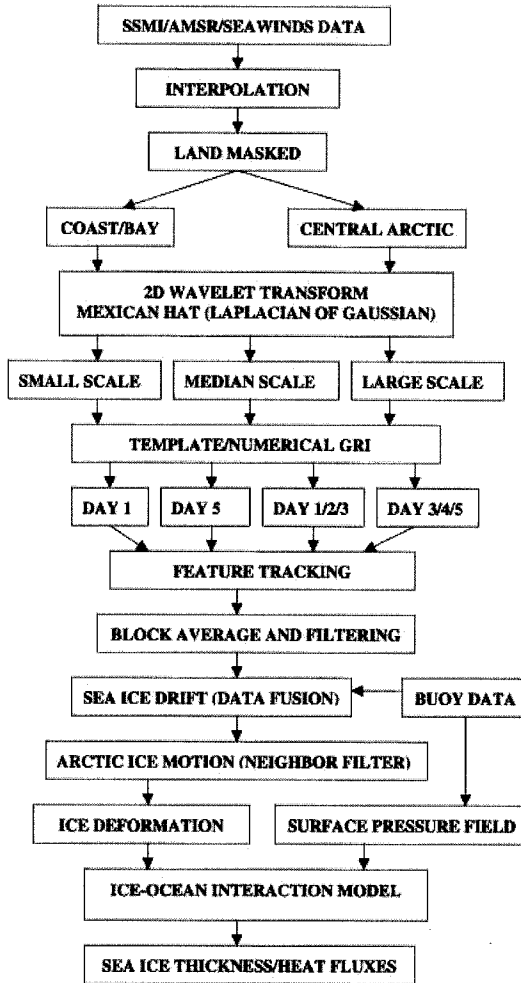


Figure 6 Flow Chart of Feature Tracking Procedure Using Wavelet Analysis of Satellite Data from SSM/I and QuikSCAT

tracking regions are considered: the coastal/bay regions for fast ice motion (with a two-day sliding window), and central Arctic for slow ice motion (with a four-day sliding window). Because the displacement of ice feature within the separation days is not expected to be very large, the domain of the template matching is restricted to an area that completely encompasses the target template at the center and is a few pixels wider than the template on all sides. To perform template matching, the template is shifted over each pixel in the domain. The summation of the absolute differences of the corresponding pixel values in the shifted template and the target template is computed for all possible locations of the shifted template. The sequence of the summation values is then used to determine the displacement vector which points to the shifted template that assumes the

minimum value of the summation. The displacement vectors whose length exceeds a pre-determined limit based on the prior knowledge of the ice motion in Arctic region are subsequently discarded. The displacement vectors obtained in this way at various spatial scales are then block-averaged to form a daily map of sea ice motion with outliers filtered out. Finally the displacement vectors are divided by the separation time to obtain the velocity vectors for the target day.

Figure 7a and b are maps of Arctic ice motion on a grid of  $100 \times 100$  km derived from wavelet analysis of SSM/I and QuikSCAT, respectively, for Dec. 13, 1999. White arrows in the figure indicate the ice velocities derived from SSM/I or QuikSCAT data, while the thick white arrows indicate the ice velocities derived from buoy data. Notice that the ice circulation has been clearly derived and that the velocities derived from satellite data agrees well with those derived from buoy data. Also, the flow patterns in these two images are extremely similar and they complement each other well. Incidentally, the areas with no velocity vectors are those where, during template matching, a matching pair to the target template cannot be found within a preset threshold. The results from SSM/I, QuikSCAT, and buoy data can be further merged to form more complete composite maps by some data fusion techniques. The following merging strategy is applied to derive merged sea ice motion. For locations where sea ice velocities from all three data sources exist, we use weighted average with the weights 0.25, 0.25, 0.5 assigned to those derived from buoy, SSM/I, and QuikSCAT data, respectively. For locations where there is no buoy observation, equal weights of 0.5 are assigned to sea ice velocities from SSM/I and QuikSCAT data sets if both exist, otherwise the weight of 1.0 is assigned to the result from the only data set that exists. Figure 8 is a merged ice motion map for December 13, 1999 that derived from SSM/I, QuikSCAT and buoy data. It is evident from Figures 7 and 8 that the merged ice motion map provides a more complete picture than the ice motion map derived from either data source alone.

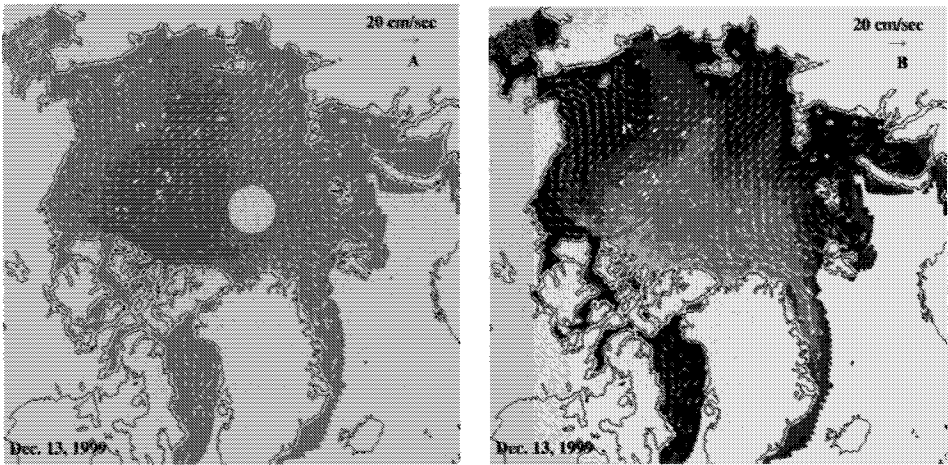


Figure 7 Arctic Sea-Ice Motion Maps in a Grid of 100 x 100 km Derived from SSM/I, and (b) QuikSCAT on December 13, 1999

## 5.2 Surface Layer Drift from MODIS and SeaWiFS

At present, there are two major global ocean color sensors in orbit: SeaWiFS and MODIS. Both satellites are in polar orbits at 705 km altitudes, and each sensor views greater than 90% of the Earth's surface in 1-2 days. SeaWiFS acquires data in 8 visible and near-infrared bands, and MODIS acquires data in 36 spectral bands (UV-VIS-IR). Several of MODIS visual wavelength bands are analogous to the SeaWiFS ocean color bands. In this case study, how images of chlorophyll *a* concentration acquired by MODIS and SeaWiFS can be used to derive surface layer drift has been demonstrated. Data from MODIS and SeaWiFS (Figure 9) were collected on May 8, 2000 at 15:45 and 16:52 GMT, respectively, off the east coast of the United States. Major oceanographic features, such as the Gulf Stream boundary and a large cold-core eddy south of the Gulf Stream, can be clearly identified.

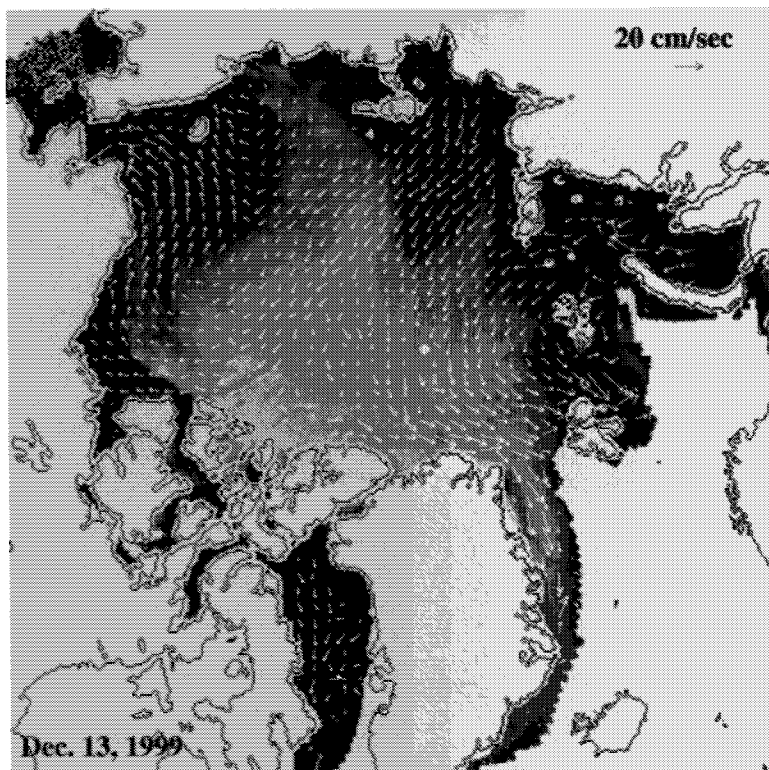


Figure 8 Merged Sea-Ice Motion Map in the Arctic on December 13, 1999

When using multiple ocean color sensors to track color feature motion, the first step is to transform the images to the same map projection. In this case, the MODIS and SeaWiFS images used are both  $1024 \times 1024$  pixels (with 1.1 km resolution), and the Mexican-hat wavelet transform is applied to filter each image with several length scales. The length scale of the wavelet transform corresponds to the length scale of the Gaussian function and is based on the length scale of the feature of interest. Filtered images, acquired 67 minutes apart, are then examined to find matching features using templates, which are then readily converted to motion vectors and averaged onto a  $17 \text{ km} \times 17 \text{ km}$  grid. The choice of 17 km as the matching template size corresponds to 16 pixels of satellite data. 17 km represents an optimal scale in a tradeoff between higher resolution and better feature tracking abilities. The accuracy of this technique is only limited by the persistence of the features and by the spatial resolution and navigational accuracy of satellite data.

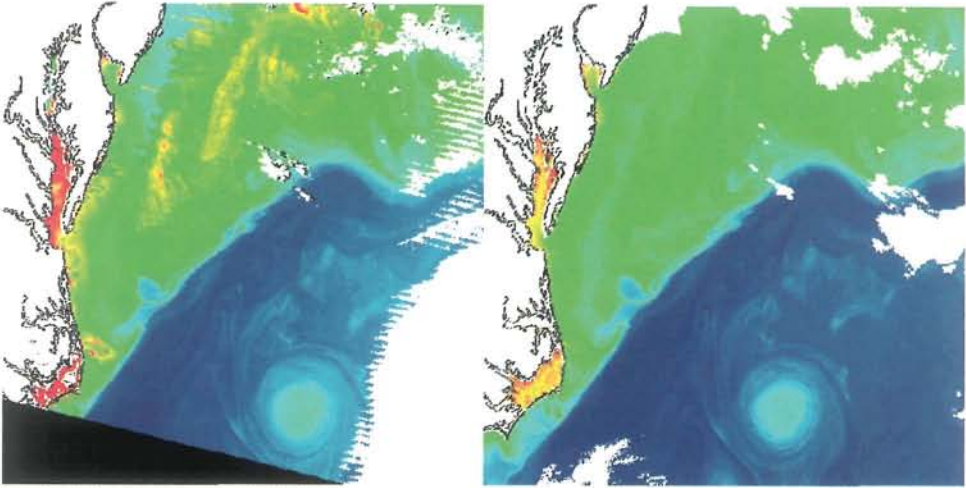


Figure 9 The Chlorophyll-a Concentration Data of Ocean Color Images Collected over the U.S. East Coast from (a) MODIS, and (b) SeaWiFS on May 8, 2000, Separated by 67 Minutes

Figure 10 shows the surface layer drift (red arrows) derived by wavelet analysis of feature tracking from MODIS and SeaWiFS chlorophyll *a* concentration data. A re-scaled MODIS image is laid as background to highlight eddies. The data were collected on May 8, 2000 over the mid-North Atlantic Ocean, separated by 67 minutes. The drifter data are shown as dark arrows in the figure and show a generally consistent pattern as compared with satellite results. As shown in this figure, both Gulf Stream current (1.5 m/s) and spin-off eddies have been well derived and can be clearly identified. Notice that the cold ring south of the Gulf Stream (center bottom of the map) is cyclonic, while the warm ring north of the Gulf Stream (right center) is anti-cyclonic. The convergence zone at Georges Bank is also clearly observable. The areas lacking drift vectors in the map indicate the regions where matching pairs cannot be found. These results indicate that same-day multiple ocean color satellite images can be used to derive surface layer drift, and can help to identify oceanic processes such as currents and eddies.

## 6. Discussion

Remote sensing with repeated coverage is the most efficient method to monitor and study marine productivity and pollution. The mapping of mesoscale ocean features in the coastal zone is a potentially major application for satellite SAR data, especially for the wide-swath SAR, such as ScanSAR (with 500 km swath) aboard RADARSAT. The use of SAR-derived observations to track eddies, surface temperature-related features, and river and estuarine plumes can aid in the management of fisheries. It is especially so in places like the Alaskan coastal area,

where uniformly cold sea surface temperatures and frequent cloud cover preclude AVHRR measurements of surface temperature features, and obscure ocean color observations.

The wavelet transforms of satellite images can be used for near real-time "quick look" screening of satellite data (feature detection), data reduction (binary image), and image enhancement (edge linking). By combining ocean color (SeaWiFS, MODIS), SAR and infrared (AVHRR) images using some data fusion techniques, mesoscale features of various physical processes such as oil spills, surface slicks, fronts, upwelling, and eddies can be detected and tracked in the coastal zone. Wavelet analysis can provide a more cost-effective monitoring program that would keep track of changes in important elements of the coastal watch system. In this study, we have demonstrated that the two-dimensional Gaussian-based wavelet analysis is a very useful tool for data fusion from multi-sensors, such as SAR, AVHRR, and SeaWiFS.

For SAR imagery, an algorithm to detect, extract, and classify linear ocean features is conceptualized and developed for automated screening, though not yet at operational level at present time. The methodology includes using histogram screening for feature detection, wavelet analysis for feature extraction, and texture analysis for feature classification. The intensity histogram of SAR images can be used for near real-time "quick look" screening of satellite data feature detection). The two-dimensional wavelet transform of a SAR image for the extraction of boundary features can be used as an edge detector with multi-scales for different type of features. More case studies of application of this algorithm to linear ocean features, such as fronts, ice edge and polar low imaged by SAR instruments, are provided by [4].



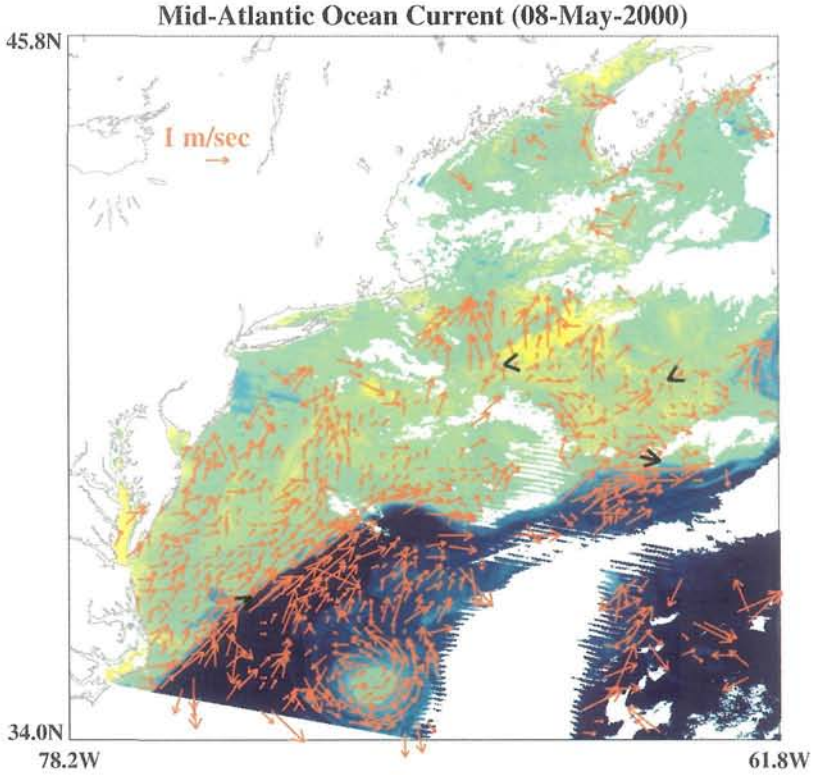


Figure 10 Surface Layer Drift (Red Arrow) Derived from MODIS and SeaWiFS Data over Mid-North Atlantic Ocean

For the sea ice motion, the technique used here provides improved spatial coverage over the existing array of Arctic Ocean buoys and better temporal resolution over techniques utilizing data from spaceborne SAR. This new source of ice motion data offers a potential solution to the problem of inadequate temporal sampling. It is found that the results from both QuikSCAT and SSM/I agree well quantitatively with ice drift derived from buoy data, and that the flow patterns illustrated by both results are extremely similar, suggesting that they would complement each other well. Furthermore, the ice motion derived from SSM/I, QuikSCAT and buoy data can be merged into one by some data fusion techniques to form composite maps that offer better coverage than results from any data source alone. Moreover, it is found that the streamlines of sea ice flow based on the merged drift data are highly correlated with the surface pressure contours, implying that the sea ice in this region is primarily driven by wind. The daily sea ice motion map can no doubt add detailed information and lend further insight into sea ice dynamics and processes through data assimilation of ocean-ice numerical model.



Analysis and tracking of ocean surface feature from low earth orbits have historically been done based on data from a single orbital sensor collected over the revisit interval of a single satellite. In the previous section, ocean surface layer drift has been derived by feature tracking of ocean color data from multiple sensors aboard different satellites. Since satellite ocean color signal comprises information from a deeper water depth (10 to 30 meters) as opposed to that from a very thin surface layer (such as sea surface temperature), one can use satellite ocean color data as a tracer for estimating surface layer currents. The example presented in Section 5.2 demonstrates how chlorophyll *a* concentration images collected by MODIS and SeaWiFS are used to derive surface layer drift. All major oceanographic features, such as Gulf Stream boundary and a large cold-core eddy south of the Stream, can be identified in the resulting drift field, confirming that the methodology is indeed a viable way to derive surface layer drift. These results of feature tracking from these multiple sensors demonstrate that wavelet analysis of satellite data is a very useful tool for satellite image processing.

### Acknowledgements

The authors would like to thank Canadian Space Agency (CSA), which owns the copyright of RADARSAT data, for providing SAR images. The QuikSCAT data is provided via the PO.DAAC at Jet Propulsion Laboratory. SSM/I data were acquired from the National Snow and Ice Data Center. SeaWiFS and MODIS data were obtained from Goddard DAAC. This research was supported through fundings by the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA).

### References

1. Combes, J.M., Grossmann, A., and Ph. Tchamitchian, 1989, "Wavelet: Time Frequency Methods and Phase Space," *Proceedings of the International Conference*, 331pp, Springer-Verlag.
2. Liu, A. K., C. Y. Peng, and S. Y.-S. Chang, 1997, "Wavelet Analysis of Satellite Images for Coastal Watch," *IEEE Journal of Oceanic Engineering*, Vol. 22, No. 1, pp. 9-17.
3. Liu, A. K., S. Martin, and R. Kwok, 1997, "Tracking of Ice Edge and Ice Floes by Wavelet Analysis of SAR Images," *Journal of Atmospheric and Oceanic Technology*, Vol. 14, pp. 1187-1198.
4. Wu, S. Y., and A. K. Liu, 2002, "An Automated Ocean Feature Detection, Extraction and Classification Algorithm for SAR Imagery," *International Journal Remote Sensing* (in press).
5. Liu, A. K., and D. J. Cavalieri, 1998, "Sea-Ice Drift from Wavelet Analysis of DMSP SSM/I Data," *International Journal of Remote Sensing*, Vol. 19, pp. 1415-1423.

6. Kwok R., A. Schweiger, D. A. Rothrock, S. Pang, and C. Kottmeier, 1998, "Sea Ice Motion from Satellite Passive Microwave Imagery Assessed with ERS SAR and Buoy Motion," *J. Geophys. Res.*, Vol. 103, pp. 8191-8214.
7. Liu, A. K., Y. Zhao, W. E. Esaias, J. W. Campbell, and T. Moore, 2002, "Surface Layer Drift Revealed by Satellite Data," *EOS, Transactions, AGU*, Vol. 83, pp. 61-64.
8. Liu, A. K., and S. Y. Wu, 2001, "Satellite Remote Sensing: SAR," *Encyclopedia of Ocean Sciences*, London: Academic Press, Edited by J. H. Steele, S. A. Thorpe, and K.K Turekian, pp. 2563-2573.
9. Liu, A. K., C. Y. Peng, and T. J. Weingartner, 1994, "Ocean-Ice Interaction in the Marginal Ice Zone Using SAR," *J. Geophys. Res.*, 99, 22391-22400.
10. Liu, A. K., Y. Zhao, and W. T. Liu, 1998, "Sea-Ice Motion Derived from Satellite Agrees with Buoy Observations," *EOS, Transactions, AGU*, Vol. 79, No. 30, pp. 353-359.
11. Liu, A. K., Y. Zhao, and S. Y. Wu, 1999, "Arctic Sea Ice Drift from Wavelet Analysis of NSCAT and SSM/I Data," *J. Geophys. Res.*, Vol. 104, pp. 11529-11538.
12. Zhao, Y., and A. K. Liu, 2002, "Validation of Sea Ice Motion from QuikSCAT with Those from SSM/I and Buoy," *IEEE Trans. Geosci. Remote Sens.* Vol. 40, pp. 1241-1246.

## CHAPTER 8

### WAVELET-BASED SAR SPECKLE FILTERS

Hua Xie, Leland Pierce, Fawwaz Ulaby  
*EECS Dept., The University of Michigan*  
*Ann Arbor, MI, 48109, USA*  
*E-mail: lep@umich.edu*

Wavelet filters have been used for despeckling of SAR images using a variety of different methods. This chapter discusses many of those previous methods and provides a framework for better understanding them and their relationship to each other. We also present 2 different techniques that use spatial correlation in different ways to perform the filtering. These techniques and others like them show promise in solving the problem of despeckling of SAR images quickly and accurately under a variety of different imaging scenarios.

#### 1. Introduction

Image speckle is an inherent property of all coherent imaging systems, including Synthetic Aperture Radar. In a SAR image, speckle manifests itself in the form of a random pixel-to-pixel variation with statistical properties similar to those of thermal noise. Due to its granular appearance in an image, speckle noise makes it very difficult to visually and automatically interpret SAR data. Therefore, speckle filtering is a critical pre-processing step for many SAR image post-processing tasks, such as segmentation and classification.

The objective of SAR speckle filtering is to reconstruct the backscattering coefficient from the image intensity. A good speckle filter should possess the following properties [40]:

- Speckle reduction in statistically homogeneous areas
- Feature preservation
- Radiometric preservation

In the past ten years, many algorithms have been developed to suppress speckle noise in order to facilitate postprocessing tasks. Two types of approaches are traditionally used. The first, often referred to as multi-look processing, involves the incoherent averaging of  $L$  single looks during the generation of the SAR image. The averaging process narrows down the probability density function (pdf) of speckle and reduces the variance by a factor  $L$ , but this is achieved at the expense of the spatial resolution (pixel area is increased by a factor  $L$ ). If the looks are not independent, such as when the Doppler bandwidth of the SAR return signal is segmented into multiple overlapping subbands, one needs to define

an equivalent number of looks (ENL) [40] to describe the speckle in the resultant images. The second approach, which is applied after the formation of the multi-look SAR image, involves the use of adaptive spatial filtering through an examination of the local statistics surrounding a given pixel. To date, various spatial filters have been developed to reduce speckle without significant loss in spatial resolution.

In 1980's, spatial filters that take into account the multiplicative speckle model were developed. The best known include the Lee [27], Kuan [25] and Frost [17] filters. A common theme among them is the use of the MMSE (Minimum Mean Square Error) technique and the assumption that the local statistics are stationary inside a moving window, which is often square-shaped. The Lee and Kuan filters are similar in that for each individual pixel, the noise-free signal is retrieved by a weighted average of pixel values in the local window, based on the criteria of MMSE. The local standard deviation usually plays a role in calculating the weighting coefficients. The only difference is that the former applies the MMSE technique on the linearized multiplicative model while the latter uses the multiplicative model directly. The Frost filter differs from the Lee and Kuan filters since the observed image is modeled as the convolution of the product of the scene reflectivity and speckle noise with the impulse response of the SAR system. Under the assumption that the radar system response is constant over some finite bandwidth and the scene reflectivity is an autoregressive process, the impulse response of the Frost filter is derived as a circular symmetric filter by using the MMSE criteria.

Lopes, *et al.* [33] introduced some practical criteria to enhance the filters in terms of speckle reduction and texture preservation. The authors proposed to divide the image under analysis into three categories using two thresholds. The first category is homogeneous areas in which a simple average filter should be applied; the second category is heterogeneous areas in which the spatial statistical filters, such as Lee, Kuan, and Frost should be utilized; the last category is areas containing an isolated point target, in which the observed value is preserved.

The effectiveness of a spatial filter relies heavily on its method of reliably measuring local statistics (for example, mean and standard deviation) through a moving window. Despite the computational simplicity of these early filters, they fail to accurately capture the underlying structure due to their fixed window size and shape. Often they cause serious degradation of features in their reconstruction. With the aim of reducing speckle in homogeneous areas while preserving the texture information in heterogeneous areas, adaptive speckle filters have been developed by adapting the size and the shape of windows according to the underlying structural features. An early example of a spatially adaptive filter is the refined Lee filter [26], an improved version of the original Lee filter in the sense that the former can achieve better spatial adaptability by incorporating local gradient information. The local relative standard deviation is calculated in a  $7 \times 7$  sliding window. After comparison with a predefined threshold, it is then determined whether an edge exists in this window. Depending on the gradient information, one of the eight pre-defined templates is chosen to calculate the weighted average in the same manner as the original Lee filter. If no edge is detected, the simple average over the whole window is assigned to the pixel of

interest.

For a spatial filter, the choice of window size and shape to use for estimating local statistics is obviously a compromise. On one hand, large windows should be chosen to minimize the influence of speckle on the estimates of the local statistics. On the other hand, large windows often contain pixels belonging to regions with different reflectivity properties, consequently, small windows are favored for better representation of fine features under such conditions. Therefore, the speckle filter with a fixed window size and shape is incapable of recovering the signal reliably. To overcome this problem, the EPOS (Edge Preserving Optimized Speckle) filter [22] was proposed to deal with the local geometry more adaptively. The shape and size of its moving window can be adjusted in a multi-resolution fashion in order to determine the largest homogeneous area around the central pixel, which is suitable for calculating the average. The EPOS filter starts with an initial large window ( $11 \times 11$ ) consisting of eight subwindows, if, based on some criteria, a homogeneous area is not detected through any combination of those subwindows, it moves to a smaller window. The window size can be continuously decreased until the smallest window ( $1 \times 1$ ) is reached. Because only averaging is carried out, the problem with the EPOS filter is its ineffectiveness in dealing with textured areas [47]. Another example of filters with variable window size is the MHR (maximum homogeneous region) filter presented in [52] by Wu and Maitre. The basic idea behind the MHR filter is very similar to the EPOS filter, aiming to find the largest homogeneous window surrounding the pixel of interest. However, the window evolution of the MHR filter is different than in the EPOS filter since the region growing technique in image segmentation is applied. The MHR is initiated with a small window ( $3 \times 3$ ). At each step, the window size is increased until the largest homogeneous window is determined, in which the original Kuan filter will be used for speckle filtering. An adaptive windowing technique also appeared in [41].

All of these filters make use of the multiplicative noise model, as well as the related speckle statistics, but they do not account for the statistics of the underlying radar clutter. If prior knowledge about the RCS (Radar Cross Section) in the product model is available, speckle filtering becomes equivalent to a Bayesian reconstruction problem. When combined with the probability distribution of speckle, the pdf of the RCS model can be used to seek the MAP (*maximum a posteriori*) estimate. A few studies [23] have reported that the Gamma distribution provides a good fit to terrain clutter. The MAP reconstruction using the Gamma distribution as the prior pdf is called the Gamma-Gamma MAP filter. Sometimes, the clutter is simply assumed to obey a Gaussian pdf, in which case the MAP filter becomes Gamma-Gaussian MAP [40] filter. Since the specific prior knowledge is incorporated into the denoising procedure, we expect such filters to yield more reliable reconstructions. Without a doubt, spatial adaptability is essential for this type of filter, because the parameters that specify the distribution need to be estimated locally.

The filters we discussed above have been the mainstream in the SAR community. It should be noted that there have been some other spatial filters, such as: [29, 41, 50, 49, 51]. With the advent of multi-frequency and fully polarimetric SAR data, the correlation of terrain reflectance between the multiple bands and multiple polarizations opens a new win-

dow to speckle reduction. It is assumed that multi-channel filters that explore the correlation between different channels will lead to a better estimate of the underlying signal than single-channel filters by which different channels are despeckled independently. The correlation is incorporated into the denoising procedure via the polarization covariance matrix or the covariance matrix in general. Representatives in the category of multi-channel speckle filters include the vector LMMSE filter developed by Lin and Allebach [31], the polarimetric whitening filter (PWF) proposed by Novak and Burl [38], and the polarimetric Lee filter [28]. The vector LMMSE filter is a generalization of the scalar LMMSE filter. The PWF constructs an optimal intensity image that has the minimum possible amount of speckle from the three complex measurements HH, HV and VV (assume HV and VH identical). While the PWF is intended to combine the complex HH, HV and VV images into one intensity image, the polarimetric Lee filter is able to handle multi-look SAR images in both intensity and amplitude formats, and produce speckle-reduced HH, VV and HV images separately. We will limit our treatment to the development of single-channel filters throughout this chapter.

A number of studies reported that spatial adaptive filters are preferable options over those that lack spatial adaptability. In principle, effective speckle reduction demands filtering algorithms capable of operation on different scales. As mentioned previously, the wavelet transform has independently arisen as a powerful tool for interpreting and processing images at different scales. More recently, a new approach has emerged in which speckle reduction is implemented in the framework of wavelet transforms. The first effort aimed at developing wavelet-based speckle filters was conducted at Rice University [14]. Guo *et al.* [21] extended Donoho's wavelet soft and hard thresholding techniques to speckle reduction for SAR imagery. Due to the fact that Donoho's algorithm was originally developed for additive white Gaussian noise (AWGN), the logarithmic transformation is applied to the SAR imagery first and the transformed speckle noise is assumed as AWGN. A number of wavelet despeckling filters have emerged following the basic approach of [21]. Gagnon and Jouan [19] considered wavelet despeckling using the complex Daubechies wavelets for the property of symmetry. In [8], a hidden Markov tree (HMT) model is proposed to describe the statistical properties of wavelet coefficients. This has given rise to the work by Sveinsson [46] where speckle reduction is studied based on the HMT. The problem of interest in [56] is speckle reduction for echocardiographic images. The authors regulated soft thresholding by applying scale-dependent thresholds, instead of a single global threshold initially proposed by Donoho. The main contributions of the literature cited above lie in transplanting the latest developments in wavelet transforms as well as in wavelet denoising for additive noise to a multiplicative speckle model. However they have a common drawback - the lack of efforts to fully characterize the statistical properties of the log-transformed speckle noise. There are other despeckling approaches reported in the literature based on wavelet multi-resolution decompositions where the log transform is not necessary. The most notable is the work done by Foucher *et al.* [16]. The authors are credited with the first attempt of MAP-based wavelet despeckling using a realistic prior model. The algorithm can be viewed as an extension of the Gamma-MAP filter to the wavelet do-

main. In [12], a speckle filtering algorithm for edge sharpening is proposed. In a nutshell, this algorithm behaves like a mean filter if no edge crossing is detected in a local window, otherwise it implements spatially adaptive averaging under the guidance of the location of the edge crossings. The derivation of this algorithm does not account for the speckle model.

The next section details the standard wavelet denoising techniques for SAR images, while section 1.3 presents our spatially-adaptive extension of these techniques. Section 1.4 then introduces another technique that avoids the use of the log operator.

## 2. Standard Wavelet Denoising

To estimate the speckle-free radar response, most spatial filters use local statistics, which are usually measured within a local sliding window. A filter's performance heavily depends on the choice of the window size and orientation. As stated by Marr [35], and also noted by other observers, "the spatial organization of a surface's reflectance function is often generated by a number of different processes, each operating at a different scale". As a result, features present in SAR imagery often exhibit different scales. This requires an adjustable local window to adapt to local spatial variations, including feature scale and geometric structure. Most filters we reviewed in section one fail to achieve spatial adaptation because they only deploy a local window with fixed size and shape. There exist a few filters which are capable of adapting the size or the shape of the local window to some extent according to the underlying structural features. The refined Lee filter [26] is such an example.

Improvement of speckle filtering can be expected with accurate and efficient feature-detection. Wavelet multiresolution analysis provides great promise for image feature detection at different scales. Wavelet decomposition has the very useful property of space and scale localization. Multiresolution analysis decomposes the original image into a set of approximations and wavelet subimages by filtering and subsampling. In view of the many theoretical developments that occurred in the last decade, wavelets have found successful applications in a variety of signal processing problems, including image compression, image denoising and image coding.

In what follows, we are concerned with the development of a SAR speckle filter in the context of wavelet denoising. We use the logarithmic transform to convert the multiplicative speckle model to an additive noise model, and assume that the resultant noise approximately follows the Gaussian distribution. The combination of the mean and the standard deviation derived from the statistical models presented in [54] will provide a good characterization for the log-transformed speckle noise. As detailed in that paper, the mean of the log-transformed speckle noise is not zero, whereas most wavelet-based denoising techniques assume AWGN with zero mean. In the literature, a number of wavelet-based despeckling algorithms [19, 21, 46] fail to address this mean-bias problem. For the purpose of radiometric preservation, the mean bias should be corrected before the exponential operation, especially for SAR images with a high noise level. In this section we present an overview of wavelet denoising algorithms. First we introduce the statistical model used in this study to characterize wavelet coefficients of natural images. Next, the hyperparameter estimation problem is addressed using the EM (Expectation and Maximization) algorithm,

followed by an implementation of the Bayesian-Markov denoising technique. Finally we outline our implementation of the translation-invariant wavelet transform, and then evaluate this new filter's performance.

## 2.1. Wavelet Denoising Overview

The wavelet transform (for example: [32, 36]) applies 4 different filters to the image, resulting in one image that is similar to a "low-pass" version of the image, with three other images, each like a "high-pass" version, but with the edge detection in different directions. These resultant images are then subsampled by two. Each of these images is called a sub-band. These filters can then be applied again, to the "low-pass" resultant image to produce four more subsampled images. Each time these filters are applied we produce images at a different scale, and each scale is called a "level" in what follows.

The essence of denoising using wavelet analysis is to reduce the noise in the wavelet transform domain. Suppose we have a length- $N$  noisy observation  $\mathbf{w} = [w_1, w_2, \dots, w_N]$ :

$$\mathbf{w} = \mathbf{f} + \epsilon, \quad (1)$$

where  $\mathbf{f} = [f_1, f_2, \dots, f_N]$  is the desired noise-free signal, and  $\epsilon = [\epsilon_1, \epsilon_2, \dots, \epsilon_N]$  is AWGN with zero mean and standard deviation  $\sigma_n$ . Because a discrete wavelet transform is a linear operator it yields an additive noise model in the transform domain:

$$\mathbf{y} = \text{DWT}(\mathbf{w}) = \text{DWT}(\mathbf{f}) + \text{DWT}(\epsilon) = \mathbf{x} + \mathbf{n}, \quad (2)$$

where DWT stands for the discrete wavelet transform and IDWT will be used to denote the inverse transform.

In order to simplify notation, here we use the 1-D vector format with boldfaced letters to represent 2-D images instead of the matrix representation. For the  $i$ th wavelet coefficient at level  $j$  in detail subband  $d$  ( $d = 1, HL$ ;  $d = 2, LH$ ;  $d = 3, HH$ ), the observation model in the wavelet domain is formulated more specifically by

$$y_{i,j}^d = x_{i,j}^d + n_{i,j}^d. \quad (3)$$

For clarity of notation, we will omit the level index  $j$  and the detail subband index  $d$  unless they are explicitly needed.

Because the wavelet scaling filter  $\{g(k)\}$  (lowpass), and the wavelet filter  $\{h(k)\}$ , (highpass) satisfy the orthogonality conditions, the resultant noise in the wavelet domain  $\mathbf{n}$  has the same statistical properties as the untransformed noise  $\epsilon$ ; i.e.,  $\mathbf{n}$  is also white Gaussian noise with zero mean and standard deviation  $\sigma_n$ .

The main scheme for recovering  $\mathbf{f}$  from  $\mathbf{w}$  using wavelet denoising techniques is summarized in the following three steps, as shown in Fig. 1:

- (1) Perform a 2-D Wavelet transform:  $y = \text{DWT}(w)$
- (2) Manipulate the wavelet coefficients:  $\hat{x} = f(y, \sigma_n)$
- (3) Perform a 2-D wavelet inverse transform:  $\hat{f} = \text{IDWT}(\hat{x})$



The main scheme for recovering  $\mathbf{f}$  from  $\mathbf{w}$  using the wavelet transform can be summarized by the three primary steps shown in Fig. 1(a).

For SAR images, a typical model used is:

$$\mathbf{w} = \mathbf{f}\epsilon. \quad (4)$$

Applying the log operator to both sides results in:

$$\log \mathbf{w} = \log \mathbf{f} + \log \epsilon. \quad (5)$$

which is now in the form of 1. This is why we interject the “log” and “exp” steps, as shown in Fig. 1(b). In the majority of wavelet denoising algorithms, noise reduction is accomplished in detail subbands with the approximation subband intact. In such cases, it is important to correct the biased mean in the approximation band introduced by the logarithmic transform. To overcome this problem, we add an additional step “adjust mean” between the “IDWT” (Inverse Discrete Wavelet Transform) and the final “exp” step. The biased mean is corrected by subtracting the mean value of the log-transformed speckle from the output image of “IDWT”.

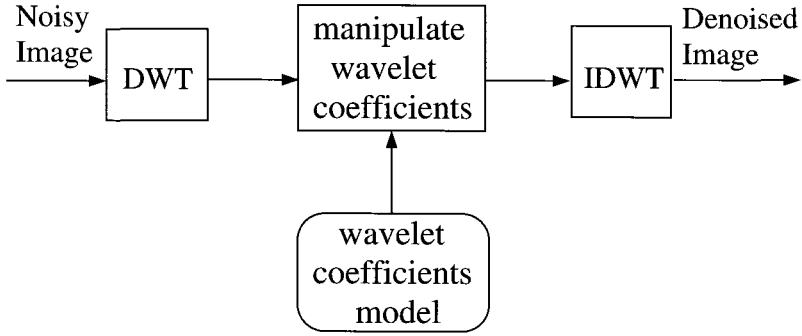
In general, the step where the wavelet coefficients are manipulated is the most crucial. What distinguishes one denoising method from another is mainly related to the approach used in this particular step. Two major denoising techniques used in this context are the thresholding technique and the Bayesian estimation shrinkage technique. In these two techniques, algorithms can be further categorized by how the wavelet coefficients are statistically modeled. Most early models [4, 6] assumed the wavelet coefficients to be independently distributed.

Later, researchers proposed more complicated, but also more accurate, models that exploit inter-scale dependencies [8], intra-scale dependencies [3, 34, 37] and the hybrid inter- and intra-scale dependencies [5, 44] among wavelet coefficients. We will discuss some of these algorithms briefly next.

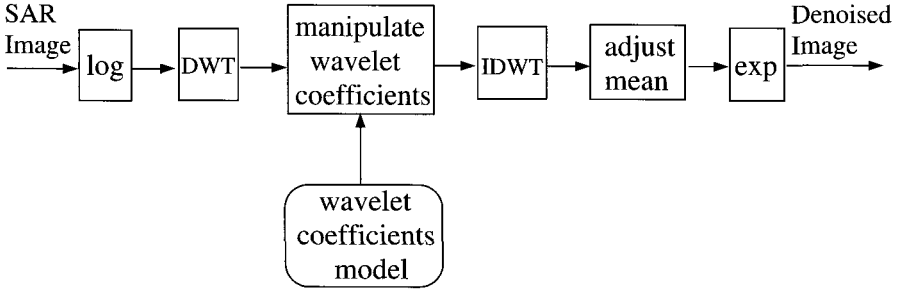
### 2.1.1. Thresholding Technique

Denoising based on thresholding in the wavelet domain was initially proposed by Donoho and Johnstone [13, 14]. Thresholding typically involves a binary decision. In [14], the authors introduced two thresholding methods, namely soft and hard thresholding. For each wavelet coefficient, if its amplitude is smaller than a predefined threshold, it will be set to zero, otherwise it will be kept unchanged (hard thresholding), or shrunk in the absolute value by an amount of the threshold (soft thresholding).

The key decision in the thresholding technique is the selection of an appropriate threshold. If this value is too small, the recovered image will remain noisy. On the other hand, if the value is too large, important image details will be smoothed out. Using a minimax criterion, Donoho proposed what the wavelet community calls the universal threshold  $T = \sqrt{2 \log(N)}\sigma_n$ , where  $N$  is the sample size and  $\sigma_n$  is the noise standard deviation. The universal thresholding technique has been recognized as simple and efficient, but if only a single threshold is used globally, it provides no spatial adaptation during the process of



(a) Additive noise model.



(b) Multiplicative speckle noise model.

Fig. 1. The wavelet denoising procedure. The filtering methods modify the wavelet coefficients in different ways. The SAR image noise model is converted from multiplicative to additive by the log operation, and un-done by adjusting the mean, and then applying the exp operator. Figure is from [53], ©2002 IEEE, used with permission.

noise suppression. Following Donoho, some researchers have focused on developing spatially adaptive thresholding techniques instead of using a global uniform threshold. In [56], a simple scaling factor function was proposed to regulate thresholds for the purpose of scale adaptation. Chang [3] first proposed a multiple threshold denoising scheme to take into account local spatial characteristics. Using the Gaussian distribution and Laplacian distribution to model wavelet coefficients, Chang *et. al* [4] proposed an approximate MMSE solution to soft-thresholding.

### 2.1.2. Bayesian Estimation

As far as Bayesian estimation is concerned, it is necessary to assume an *a priori* distribution  $p(\mathbf{x})$  associated with the wavelet coefficients of the noise-free image. Given the likelihood function  $p(\mathbf{y}|\mathbf{x})$ , and assuming that  $\mathbf{x}$  is independently and identically Gaussian distributed with zero mean and variance  $\sigma_x^2$ , given the AWGN model, we can estimate the noise-free wavelet coefficients  $\mathbf{x}$  by [44]:

$$\hat{x}_i = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_n^2} y_i. \quad (6)$$

The deficiencies associated with this shrinking function are twofold. First, the assumed prior disagrees with the strong non-Gaussian statistics exhibited by wavelet coefficients of natural images. Secondly, each wavelet coefficient is denoised individually with the lack of spatial adaptation toward the intra-scale and inter-scale dependencies.

In the remainder of this section, we therefore adopt the mixture of Gaussian densities [6, 8] to model natural images, meanwhile characterizing the intrascale contextual dependence of wavelet coefficients using Markov random fields (MRF). The idea of exploiting the clustering property of wavelet coefficients using MRF also appears in [34].

To capture the insignificant/significant coefficient property, for each wavelet coefficient we define a binary hidden state  $s_i$ , which can take on the value 0 (insignificant coefficient) or 1 (significant coefficient). The configuration of  $s_i$  over the entire wavelet subband image forms a binary mask  $\mathbf{s}$ . The marginal density of wavelet coefficients is defined as:

$$p(x_i) = \sum_{k=0,1} p(s_i = k) p(x_i | s_i = k), \quad (7)$$

$$p(x_i | s_i = k) \sim \mathcal{N}(0, \sigma_{xk}^2), \quad k = 0, 1 \quad (8)$$

where  $\mathcal{N}(0, \sigma_{xk}^2)$  stands for a Gaussian distribution with zero mean and standard deviation  $\sigma_{xk}$ .

When the signal is corrupted by additive white Gaussian noise, it is easy to show that the noisy wavelet coefficients also obey the mixture density of two normal distributions with zero mean, but with an increased variance that depends on the noise level. That is:

$$p(y_i) = \sum_{k=0,1} p(s_i = k) p(y_i | s_i = k), \quad (9)$$

$$p(y_i | s_i = k) \sim \mathcal{N}(0, \sigma_{xk}^2 + \sigma_n^2), \quad k = 0, 1. \quad (10)$$

Given  $\sigma_n^2$ , the statistical model is fully parameterized by three independent parameters,  $\sigma_{x1}, \sigma_{x2}$ , and  $p(s_i = 0)$ . These parameters can be grouped into a model hyperparameter vector  $\Theta$ . A total of  $3J$  sets of  $\Theta$  need to be estimated for the three orientation subbands across  $J$  decomposition levels. In [6], hyperparameters are chosen intuitively and empirically. We use the EM algorithm to estimate the hyperparameters because it is automatic and it will result in more accurate estimates.

The general EM algorithm was first formalized by Dempster, *et al.* [10] in order to obtain ML-estimates from incomplete data. Consider the observed data  $\mathbf{x}$  as the incomplete data. We assume that there exists a complete data set  $\mathbf{z} = (x, y)$  and also assume

a joint density function:  $p(\mathbf{z}|\Theta) = p(\mathbf{x}, y|\Theta)$ . Given the new density function, we define a new likelihood function  $L(\Theta|\mathbf{x}, y) = p(\mathbf{x}, y|\Theta)$ , called the complete-data likelihood, which is parameterized by  $\Theta$ . The goal of EM in its basic form is to seek the solution  $\Theta$  which maximizes the log-likelihood function  $\log p(\mathbf{x}, y|\Theta)$ , instead of  $\log p(\mathbf{x}|\Theta)$ . The algorithm starts with an initial value  $\Theta^0$ , and then it proceeds iteratively. Each EM iteration consists of an E-step and an M-step. The E step is to evaluate the probability distribution for the data given the model parameters from the previous iteration, i.e.,  $Q(\Theta|\Theta^{i-1}) = E[\log p(\mathbf{x}, y|\Theta)|\mathbf{x}, y, \Theta^{i-1}]$ , and the M step is to find  $\Theta^i$  by maximizing  $Q(\Theta|\Theta^{i-1})$ . The two steps are iterated until convergence is reached.

In this study, we have a mixture Gaussian model:

$$p(\mathbf{x}|\Theta) = \sum_{j=1}^m \alpha_j p(\mathbf{x}|\mu_j, \Sigma_j) \quad (11)$$

$$p(\mathbf{x}|\mu_j, \Sigma_j) = \frac{1}{\sqrt{2\pi}|\Sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu_j)^T \Sigma_j^{-1}(\mathbf{x}-\mu_j)} \quad (12)$$

where  $\mu_j$  and  $\Sigma_j$  are the mean value and the covariance matrix of the  $j$ th component. For the mixture model problem, we introduce the unobserved data  $\mathbf{y} = (y_1, y_2, \dots, y_N)$  which indicates the origin of the component distribution [42]. For each  $i$ ,  $y_i$  is an integer between 1 and  $m$ , and  $y_i = j$  means the  $i$ th sample originates from the  $j$ th mixture component. The parameter vector  $\Theta = (\alpha_1, \dots, \alpha_m; \mu_1, \dots, \mu_m; \Sigma_1, \dots, \Sigma_m)$  can be estimated by the EM algorithm as follows [42]:

$$\alpha_{k,j}^i = \frac{\alpha_j^{i-1} p_j(x_k|\boldsymbol{\theta}_j^{i-1})}{p(x_k|\boldsymbol{\theta}^{i-1})}, \quad \alpha_j^i = \frac{1}{N} \sum_{k=1}^N \alpha_{k,j}^i, \quad (13)$$

$$\boldsymbol{\mu}_j^i = \frac{\sum_{k=1}^N \alpha_{k,j}^i x_k}{\sum_{k=1}^N \alpha_{k,j}^i}, \quad (14)$$

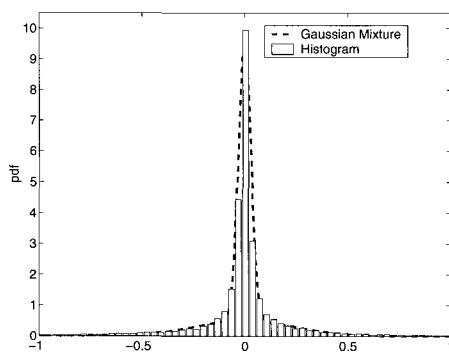
$$\Sigma_j^i = \frac{\sum_{k=1}^N \alpha_{k,j}^i (x_k - \boldsymbol{\mu}_j^i)(x_k - \boldsymbol{\mu}_j^i)^T}{\sum_{k=1}^N \alpha_{k,j}^i}. \quad (15)$$

By way of illustration, we applied the EM (Expectation and Maximization) algorithm described briefly above to fit the two-Gaussian density model to the wavelet coefficient histogram of the Lena subbands at level one, with and without speckle noise. The noisy version of the Lena image has 3-look speckle in amplitude format. Table 1 lists the estimates of the mixture Gaussian model hyperparameters. In Table 1, the fact that  $p(s=0) > p(s=1)$  confirms the sparse representation provided by the wavelet transform. The majority of the wavelet coefficients are represented by the Gaussian distribution with the low variance  $\sigma_{x0}$ , while the remaining significant coefficients obey the Gaussian distribution with the high variance  $\sigma_{x1}$ . In order to illustrate how the Gaussian mixture model fits the wavelet coefficients, we compare the model based on hyperparameters listed in Table 1 with the histogram of the noise-free and noisy *Lena HLL1* subband, in Fig. 2(a) and Fig. 3(a), respectively. For both cases, we also plot their corresponding two mixture components in

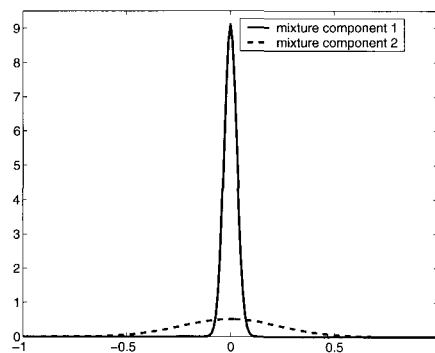
Fig. 2(b) and Fig. 3(b).

Next we explore how to use a Markov random field to incorporate spatial dependencies into the denoising procedure.

Hyper-parameters	LH1		HL1		HH1	
	x	y	x	y	x	y
$\sigma_{x0}$	0.022	0.010	0.031	0.012	0.016	0.009
$\sigma_{x1}$	0.155	0.136	0.271	0.370	0.109	0.116
$p(s = 0)$	0.62	0.71	0.63	0.68	0.56	0.70
$p(s = 1)$	0.38	0.29	0.37	0.32	0.44	0.30



(a) The Gaussian mixture density



(b) Two mixture components

Fig. 2. The Gaussian mixture density for subband  $HL_1$  of the log-transformed noise-free *Lena* image. Note that the 2 mixture components model the data very well, and that a single Gaussian density could not deal with the tails correctly. Figure is from [53], ©2002 IEEE, used with permission.

### 3. MRF Based Spatially Adaptive Bayesian Wavelet Denoising

A problem in the majority of automated information extraction techniques using remote sensing images is that they do not consider the pixel's spatial context during the process. As a result, discrimination accuracy is not always very good. In order to incorporate the rich information in the spatial constraints that the human vision system uses successfully into the process of image interpretation, the Markov random field (MRF) model is one of the solutions generally proposed.

A Markov Random Field is a stochastic process defined on a 2-D lattice system to specify its contextual characteristics. MRFs form a fundamental framework for various image modeling and processing problems [30]. In this section, we will use a Markov random field

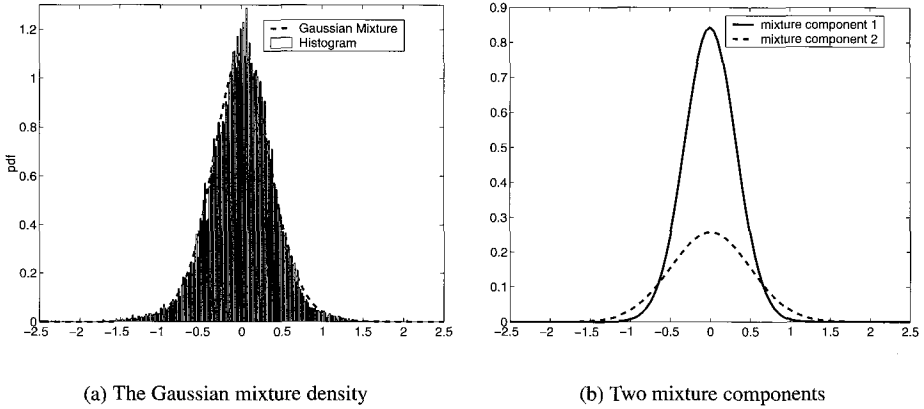


Fig. 3. The Gaussian mixture density for subband  $HL_1$  of the log-transformed noisy *Lena* image with 3-look speckle in amplitude format. Note that in this case, as compared with the previous figure, a single Gaussian density can more successfully model the data. Figure is from [53], ©2002 IEEE, used with permission.

to model the intrascale spatial dependence between wavelet coefficients in each individual subband.

After the hyperparameters specifying the mixture Gaussian density have been estimated, our proposed denoising algorithm[53] proceeds as follows:

- Calculate the shrinkage function using the Bayesian MMSE estimation technique
- Generate an initial binary mask,  $\mathbf{s}$ , corresponding to the hidden state configuration using MAP
- Redefine the prior  $p(\mathbf{s})$  using MRF, then refine the binary mask by maximizing  $p(\mathbf{s}|y)$
- Modify the shrinkage function based on the optimal binary mask  $\mathbf{s}$ .

The estimate of the noise-free wavelet coefficients is [6]:

$$\hat{x}_i = \sum_{k=0,1} p(s_i = k|y_i) \frac{\sigma_{xk}^2}{\sigma_{xk}^2 + \sigma_n^2} y_i = \xi_i y_i \quad (16)$$

where

$$\xi_i = \sum_{k=0,1} p(s_i = k|y_i) \frac{\sigma_{xk}^2}{\sigma_{xk}^2 + \sigma_n^2} \quad (17)$$

is called the shrinkage factor. Equation (16) shows that the Bayesian MMSE estimate of  $x$  is a nonlinear rescaling of the noisy observation  $y$ .

To take spatial dependence into account, we introduce a Markov random field prior model for the hidden state configuration,  $\mathbf{s}$ , of the wavelet coefficients. In order to distinguish it from the prior  $p(\mathbf{s})$  defined previously, we denote the new prior  $p_{new}(\mathbf{s})$ .

The prior is described by a two-state Potts model [30] with a second order neighborhood system. Furthermore, only single-site and pair-site cliques are considered. The prior

is defined as:

$$p_{new}(\mathbf{s}) = \frac{1}{Z} e^{-U(\mathbf{s})} \quad (18)$$

$$U(\mathbf{s}) = V_1(\mathbf{s}) + V_2(\mathbf{s}) \quad (19)$$

$$e^{-V_1(\mathbf{s})} = \prod_{i=1}^N p(s_i) \quad (20)$$

$$V_2(\mathbf{s}) = -\beta \sum_{i=1}^N \sum_{j \in \eta(i)} (\delta(s_i - s_j) - 1) \quad (21)$$

where  $Z$  is a normalizing constant,  $\beta$  is the parameter which controls the local smoothness,  $\eta(i)$  is the second-order neighborhood of pixel  $i$ ,  $\delta(\cdot)$  is the discrete delta function, and  $V_1$  and  $V_2$  are the single-site and pair-site clique functions respectively, which are defined in such a way that the former represents the prior knowledge of the distribution of the hidden state without considering spatial constraints, and the latter favors the neighboring pixels to have the same label.

With  $\mathbf{s}$  being the binary image of the hidden state, where each pixel can take on two states 0 or 1, and  $\mathbf{y}$  being the noisy wavelet coefficients, we assume that each wavelet coefficient  $y_i$  is conditionally independent of all other coefficients conditioned on knowledge of the hidden state at that location,  $s_i$ . This assumption is summarized by the following likelihood function  $p(\mathbf{y}|\mathbf{s})$ :

$$p(\mathbf{y}|\mathbf{s}) = \prod_{i=1}^N p(y_i|s_i) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi(\sigma_{x_{s_i}}^2 + \sigma_n^2)}} e^{-\frac{y_i^2}{2(\sigma_{x_{s_i}}^2 + \sigma_n^2)}}. \quad (22)$$

Combining the likelihood function and the prior model, we can seek the MAP solution of  $p(\mathbf{s}|\mathbf{y})$  as an optimal configuration:

$$\hat{\mathbf{s}} = \arg \max p(\mathbf{s}|\mathbf{y}) = \arg \max p(\mathbf{y}|\mathbf{s})p_{new}(\mathbf{s}). \quad (23)$$

In this work, maximizing the posterior probability  $p(\mathbf{s}|\mathbf{y})$  is implemented by the iterative conditional modes (ICM) algorithm, which is summarized in the next subsection. In general, the convergence rate of this optimization scheme is very much dependent on the initial condition; therefore, we generate a reasonable initial configuration of the hidden state by using the MPM (Marginal *A Posteriori* Mode) criterion based on the old independent prior. For each wavelet coefficient, its associated initial hidden state  $\hat{s}_i^0$  is determined as:

$$\hat{s}_i^0 = \arg \max_{s_i \in (0,1)} p(s_i|y_i) = \arg \max_{s_i \in (0,1)} p(y_i|s_i)p(s_i). \quad (24)$$

### 3.1. The Optimization Scheme - ICM (Iterated Conditional Modes)

Once the binary mask refinement problem has been formulated as a maximization/minimization problem, it can be solved by an optimization algorithm. Bayesian methods coupled with Markovian modeling usually result in a nonconvex cost or energy function, which could have many local maxima/minima. Global optimization algorithms, such

as simulated annealing [1] could be applied. To speed up the convergence, Besag [2] proposed a deterministic optimization algorithm called the iterated conditional modes (ICM) method which maximizes the local conditional probability iteratively.

Suppose that  $\hat{\mathbf{x}}$  is an estimate of the true  $\mathbf{x}$ , the ICM algorithm proceeds to update the current state  $\hat{x}_i$  at pixel  $i$  using the information of observation  $\mathbf{y}$  and the current configuration  $\hat{\mathbf{x}}$  at any place except pixel  $i$ . Following Bayes' theorem, we have [2]:  $p(x_i|\mathbf{y}, \hat{x}_{S \setminus i}) \propto p(y_i|x_i)p(x_i|\hat{x}_{\eta_i})$  where  $\eta_i$  is the neighborhood of pixel  $i$ , and  $S \setminus i$  denotes the set of all the pixels in the image except pixel  $i$ . The new state of  $x_i$  can be found by maximizing the local distribution  $p(x_i|\mathbf{y}, \hat{x}_{S \setminus i})$ . Note that [2]:  $p(\mathbf{x}|\mathbf{y}) = p(x_i|\mathbf{y}, x_{S \setminus i})p(x_{S \setminus i}|\mathbf{y})$ . Because the update at  $x_i$  does not affect  $p(x_{S \setminus i}|\mathbf{y})$ , it is apparent that  $p(\hat{\mathbf{x}}|\mathbf{y})$  never decreases at any stage [2]. This guarantees the eventual convergence, but only a local maximum will be achieved. The procedure completes an iteration after being applied to each site within the image. We can apply the algorithm for a fixed number of iterations, or until there is evidence of convergence [24]:

- (1) start with an initial configuration and set  $k = 0$
- (2) for each site  $i$ , update the current state  $x_i$  with  $\hat{x}_i$  which maximizes the local  $p(x_i|\mathbf{y}, x_{S \setminus i})$
- (3) go to step 2 with  $k = k + 1$  until convergence is reached

In what follows, we derive the local property that is specified through the conditional probability:

$$p(s_k|s_{l,l \neq k}) = \frac{p(s_k, s_{l,l \neq k})}{p(s_{l,l \neq k})} = \frac{p(S)}{\sum_{s_k} p(S)} \quad (25)$$

Substituting into the above equation, we have:

$$p(s_k|s_{l,l \neq k}) = \frac{\prod_{i=1}^N p(s_i) \exp(\beta \sum_{i=1}^N \sum_{j \in \eta(i)} (\delta(s_i - s_j) - 1))}{\sum_{s_k} \prod_{i=1}^N p(s_i) \exp(\beta \sum_{i=1}^N \sum_{j \in \eta(i)} (\delta(s_i - s_j) - 1))} \quad (26)$$

Divide both the numerator and denominator into two sets, one containing the site  $s_k$  and the other not containing the site  $s_k$ . After a few steps of straightforward algebraic manipulation, the above can be written as:

$$p(s_k|s_{l,l \neq k}) = p(s_k|s_{l,l \in \eta(k)}) = \frac{1}{Z'} p_{old}(s_k) \exp\left(2\beta \sum_{l \in \eta(k)} (\delta(s_k - s_l) - 1)\right) \quad (27)$$

where  $Z'$  is a constant that does not depend on  $s_k$ . Using the ICM method, for the wavelet coefficient at pixel  $i$ , its optimal hidden state  $s_i$  can be iteratively updated by:

$$\hat{s}_i = \arg \max_{s_i \in (0,1)} \left\{ -\frac{1}{2} \log(\sigma_{x_{si}}^2 + \sigma_n^2) - \frac{y_i^2}{2(\sigma_{x_{si}}^2 + \sigma_n^2)} + \log(p(s_i)) + 2\beta \sum_{j \in \eta(i)} (\delta(s_i - s_j) - 1) \right\}. \quad (28)$$



Shown in Figure 4 are the initial binary mask and its MAP estimate for the HL detail image of Lena at level 1. In both figures, bright pixels indicate the detected significant wavelet coefficients, while dark pixels represent the insignificant coefficients. It is obvious that in Figure 4(a), there exist some dubious responses due to noise in smooth regions. The optimization process led to false alarm reduction and detection improvement. As a result, we obtain a “cleaner” binary mask in Figure 4(b).

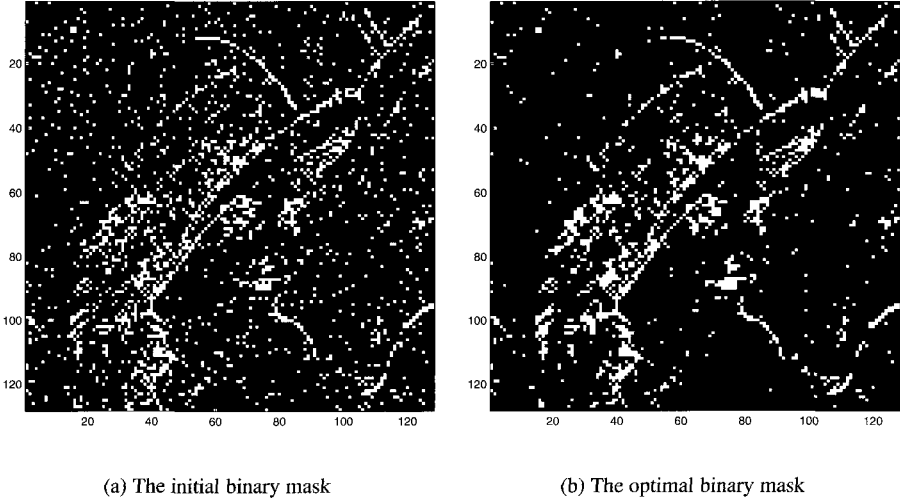


Fig. 4. The MAP estimate of the hidden state for HL detail image of Lena at level 1. Figure is from [53], ©2002 IEEE, used with permission.

Once the optimal binary mask is produced, for each wavelet coefficient, we search all its neighboring pixels with the same hidden state in its second-order neighborhood, then we calculate a weighted average of shrinkage factors of those pixels and assign it to the current pixel. To account for the directional property represented by different detail subbands, we consider a set of subband-dependent weighting coefficients for subbands  $HL$ ,  $LH$  and  $HH$ . Nevertheless, we use the same set of weighting coefficients for the same orientation subbands across the decomposition levels. Figure 5 shows the values of the weighting coefficients  $\{\omega_i\}$  when the entire  $3 \times 3$  window is occupied by only one state. Pixels with a different hidden state from the central pixel will be excluded from the averaging process, consequently normalization is necessary in order to preserve the intensity level. In summary, given the optimized binary mask  $\mathbf{s}$ , the shrinkage factor for the central pixel  $i$  in a  $3 \times 3$  local window is modified as

$$\hat{\xi}_i = \frac{\sum_j \omega_j \xi_j \delta(s_j - s_i)}{\sum_j \omega_j \delta(s_j - s_i)}. \quad (29)$$

We can now estimate the noise-free wavelet coefficient by  $\hat{x}_i = \hat{\xi}_i y_i$ .

1/18	1/18	1/18
1/6	1/3	1/6
1/18	1/18	1/18

1/18	1/6	1/18
1/18	1/3	1/18
1/18	1/6	1/18

1/8	1/24	1/8
1/24	1/3	1/24
1/8	1/24	1/8

Fig. 5. Weighted averaging coefficients  $\{\omega_i\}$  used in a  $3 \times 3$  window for subband  $LH$ ,  $HL$  and  $HH$ . Figure is from [53], ©2002 IEEE, used with permission.

### 3.2. Application to SAR Images

In this section, we illustrate the performance of our proposed spatially adaptive wavelet despeckling algorithm for SAR images in both intensity and amplitude formats.

In addition to the denoising techniques and the statistical models of wavelet coefficients as we discussed earlier, the choice of the underlying wavelet family as well as the length of the wavelet filter is also important for the overall denoising performance. The length of a wavelet filter is related to smoothness and localization properties. The primary advantage of short wavelet filters, such as the Haar wavelet, is their compact spatial support; in contrast, longer wavelet filters are preferred for smoothness. Furthermore, computational time of the wavelet transform is approximately proportional to the length of the wavelet filter [21]. To compromise, we employed Daubechies' maximally flat wavelet [9] of length 6 to implement the orthogonal DWT throughout this work. To allow for robust and reliable parameter estimation, we maintained the minimum dimension of a wavelet subband at  $32 \times 32$ . For the simulated SAR images of  $256 \times 256$  used in this study, the decomposition level was therefore limited to three. At each decomposition level, the denoising algorithm was applied separately to three detail subbands, while the approximation subband was left intact. For simplicity, we used  $\beta = 1.5$  for the Potts model to control local smoothness. A few papers [15], [45] have reported that  $\beta = 1.5$  usually gives satisfactory results. Other than being fixed to a constant beforehand,  $\beta$  can be estimated in a more justifiable way by applying statistical inference approaches [30]. More specifically, labeling (the optimal  $s$ ) and estimation (the model parameter  $\beta$ ) have to be performed alternately from the observation ( $y$ ). This is performed at the expense of computational complexity. In terms of the convergence rate of the ICM, Besag reported that a local maximum can be reached by six to eight cycles of the ICM [2]. To speed up the entire procedure, we optimized each binary mask using the ICM with four cycles. The termination rule for the iterative EM algorithm is a logical "or" of two conditions: one is exceeding the maximum number of iterations (30), the other is that the maximum relative change of parameters between two adjacent iterations is less than a threshold (1%).

The other methods against which we assess the performance of the proposed speckle filter include:

- The Bayesian soft thresholding technique proposed in [4] and described in Section 2.
- The Bayesian MMSE estimation technique using the Gaussian mixture density model

developed in [6] and described in this section.

- The refined Lee filter [26].

We have conducted a thorough study of speckle filter performance and have found that among commonly used spatial filters, the refined Lee filter [26] possesses the best trade-off between noise reduction and feature preservation. Therefore we also include the refined Lee filter in the comparison, with its local window being set up as a classical  $7 \times 7$ .

### Simulated SAR images

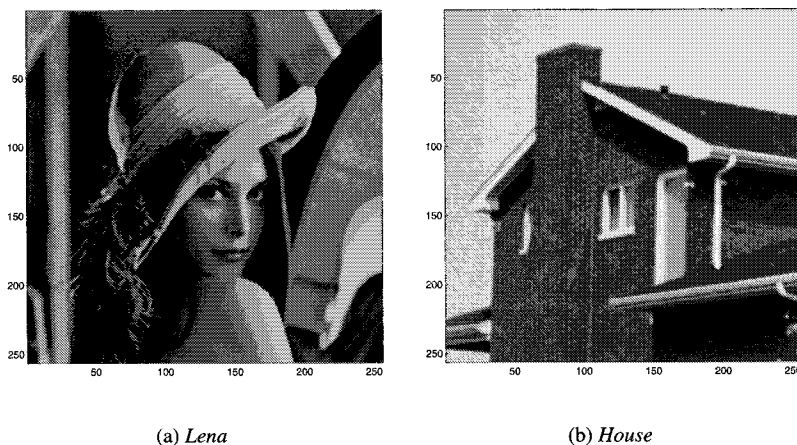


Fig. 6. Two test images used in this study. Figure is from [53], ©2002 IEEE, used with permission.

The reference images used in this study include the standard  $256 \times 256$  *Lena* and *House* images shown in Fig. 6, both of which are used frequently by the wavelet community. We corrupted them by multiplying simulated spatially uncorrelated speckle noise. In a SAR image, a number of independent pixels are averaged together to generate an image. For  $N$  independent samples this is called an  $N$ -look image, with  $N$ -look speckle statistics [20]. For quantitative evaluation, the following measures have been used:

- SNR: the signal-to-noise ratio is defined as the ratio of the variance of the noise-free signal  $f$  to the mean-squared error (MSE) between the noise-free image and the de-speckled image  $\hat{f}$  [44]:

$$\text{SNR} = 10 \log_{10} \frac{\text{var}(f)}{\frac{1}{N} \sum_{i=1}^N (f_i - \hat{f}_i)^2} \quad (30)$$

- ENL: the equivalent number of looks is defined as

$$\text{ENL} = A \frac{(\text{mean})^2}{\text{variance}}, \quad A = \begin{cases} 1, & \text{Intensity format} \\ \frac{4}{\pi} - 1, & \text{Amplitude format} \end{cases} \quad (31)$$

with the mean and the variance measured within a homogeneous region.

Using a 3-level wavelet decomposition, Tables 2-5 list the SNR values for the *Lena* image and the *House* image at nine noise levels (nine different numbers of looks) and two SAR image formats (intensity and amplitude), with the highest SNR values highlighted. It should be noted that the SNR is calculated on an average of ten independent noise samples. The experiment was repeated ten times under exactly the same settings except that speckle was realized using different random seeds, but with the same distribution.

Looks	1	2	3	9	16	25	36
Noise	-6.63	-3.62	-1.86	2.91	5.41	7.35	8.94
Baye. Thr.	<b>5.43</b>	7.39	8.41	10.83	11.97	12.96	13.82
Baye. MMSE.	5.21	7.38	8.36	11.19	12.56	13.71	14.69
This Study	5.39	<b>7.60</b>	<b>8.66</b>	<b>11.48</b>	<b>12.94</b>	<b>14.07</b>	<b>15.13</b>
Refined Lee	4.99	6.84	7.94	11.08	12.74	14.00	15.03

Looks	1	2	3	9	16	25	36
Noise	-0.99	2.02	3.78	8.55	11.05	12.99	14.57
Baye. Thr.	8.37	10.13	11.08	13.58	15.01	16.22	17.28
Baye. MMSE.	8.57	10.42	11.47	14.47	16.07	17.33	18.55
This Study	<b>8.74</b>	<b>10.65</b>	<b>11.75</b>	<b>14.82</b>	<b>16.55</b>	<b>17.84</b>	<b>18.95</b>
Refined Lee	8.55	10.48	11.62	14.76	16.47	17.78	18.90

Looks	1	2	3	9	16	25	36
Noise	-8.32	-5.31	-3.55	1.22	3.72	5.66	7.24
Baye. Thr.	<b>5.52</b>	8.27	9.34	11.99	13.20	14.18	15.00
Baye. MMSE.	5.37	8.26	9.35	12.10	13.49	14.61	15.43
This Study	5.42	<b>8.30</b>	<b>9.43</b>	<b>12.24</b>	<b>13.65</b>	<b>14.80</b>	<b>15.68</b>
Refined Lee	4.35	6.25	7.48	11.00	12.82	14.14	15.18

As we can see from these tables, the proposed method consistently outperforms the other two wavelet-based denoising algorithms in most cases ( $L \geq 2$ ). For the *Lena* image, when compared with the Bayesian MMSE estimation method, the improvement is about 0.2dB-0.5dB, depending on the noise level, for both intensity and amplitude formats. The SNR improvement over the Bayesian soft thresholding method ranges from 0.2dB to 1.3dB for the intensity format, and from 0.4dB to 1.7dB for the amplitude format. For the *House*

Looks	1	2	3	9	16	25	36
Noise	-2.68	0.33	2.09	6.86	9.36	11.29	12.88
Baye. Thr.	9.27	11.26	12.23	14.76	16.10	17.14	18.06
Baye. MMSE.	9.30	11.31	12.42	15.22	16.58	17.77	18.75
This Study	<b>9.42</b>	<b>11.41</b>	<b>12.56</b>	<b>15.37</b>	<b>16.80</b>	<b>18.05</b>	<b>19.05</b>
Refined Lee	8.20	10.36	11.60	14.92	16.63	17.89	18.97

image, our method achieved only a slight improvement, up to 0.2dB over the MMSE estimation method, and up to 0.7dB over the Bayesian soft thresholding technique for the intensity format. The improvement gets moderately better for the amplitude format. The reason why the proposed method failed to yield significant improvement is as follows. We assume a Gaussian mixture density for the wavelet coefficients of natural images. Subsequent log-compression decreases the dynamic range of the data, making it very difficult to estimate the hyperparameters that specify the mixture model for images in which texture only comprises a small fraction of the total image, which is the case for the *House* image.

When the image is corrupted by single-look speckle in intensity format, soft thresholding achieves slightly better performance than the other two wavelet-based algorithms. There are two reasons for this: first, when the noise level is very high, the log-transformed speckle noise deviates strikingly from the Gaussian distribution; and second, the EM algorithm is not able to provide accurate estimates of the three hyperparameters for the Gaussian mixture model, whereas estimation is more robust in the case of soft thresholding since only one parameter is needed.

Compared with the refined Lee filter, we observe that our method has higher SNRs at high noise levels. As the noise level decreases, they achieve almost equivalent performance with respect to SNR. In terms of the algorithm complexity, the refined Lee filter appears to be much simpler and more straightforward than the proposed method. However the denoising performance of the proposed method can be further improved by the cycle spinning technique [7].

To implement cycle spinning, a total of 64 shifts ( $8 \times 8$  shifts in row and column directions) were used, corresponding to the 3-level wavelet decomposition conducted in this study. In Fig. 3.2, the SNR values as a function of shifts using the proposed method are plotted for the *Lena* image. We find that cycle spinning can substantially improve the denoising performance. With  $2 \times 2$  shifts, SNR can be increased by up to 1dB, and a further 0.5dB increment can be achieved with  $4 \times 4$  shifts. However  $8 \times 8$  shifts do not result in significant improvement over  $4 \times 4$  shifts. In most cases, the gain is marginal, less than 0.1dB. Therefore cycle spinning with  $4 \times 4$  shifts is sufficient for a  $256 \times 256$  image.

Method	Speckled	Baye. Thr	Baye. MMSE.	This Study	Refined Lee
Region 1	16.9	155.5	221.9	287.5	244.6
Region 2	15.5	138.2	186.8	228.2	212.6

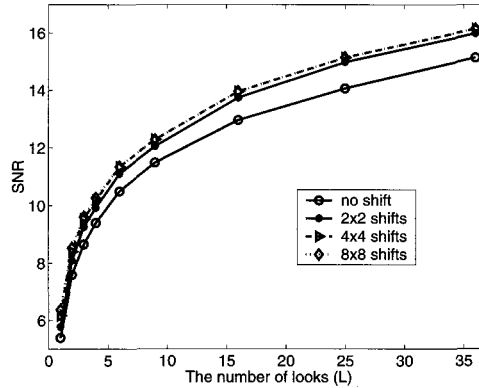


Fig. 7. SNR as a function of the number of shifts in cycle spinning for the Lena image. Figure is from [53], ©2002 IEEE, used with permission.

Fig. 8 presents a comparison of the denoised *Lena* image corrupted with speckle noise ( $L = 16$ ) in intensity format. Cycle spinning with  $4 \times 4$  shifts was implemented for the three wavelet-based denoising algorithms. It is observed that many isolated specks in smooth regions present in Fig. 8(c) disappear from Fig. 8(d) due to the MRF-based optimization procedure. Table 6 lists the equivalent number of looks estimated within two statistical homogeneous areas in the *Lena* image for different filters. It indicates that the proposed filter achieved the most successful noise reduction in homogeneous areas. On the other hand, we find that the refined Lee filter produced sharper edges, but in the homogeneous areas, due to the effect of sliding windows, artifacts are visible. Pixel profiles at column 185 are plotted in Fig. 3.2 for the original noise-free image and the filtered images by the proposed method and the refined Lee filter respectively.

Test Areas	Orig.	Baye. Thr.	Baye. MMSE	This Study	Refined Lee
Region 1	4.2	30.1	52.8	63.7	51.0
Region 2	4.2	32.3	58.4	70.3	56.3
Region 3	4.4	33.8	59.6	76.1	62.2

All the computation was carried out using programs written in C and running on a SUN Ultra 1 machine. Using a three-level DWT, the CPU time for denoising the *Lena* image is about 0.9s, 7.5s, and 8.7s for the Bayesian soft thresholding, the Bayesian MMSE, and the proposed method, respectively. The refined Lee filter requires about 2.0s. For both the Bayesian MMSE and the proposed method, it was found that a large portion of their execution time is spent on Gaussian mixture parameter estimation using the EM algorithm. When the cycle spinning technique is added, we can assume these parameters are the same for different shifts, therefore the time-consuming EM algorithm only need be applied once.



(a) Noisy observation



(b) Bayesian soft thresholding



(c) Bayesian MMSE estimation



(d) The proposed method



(e) Refined Lee

Fig. 8. Comparison of different despeckling methods for *Lena* image with  $L = 16$ . Figure is from [53], ©2002 IEEE, used with permission.

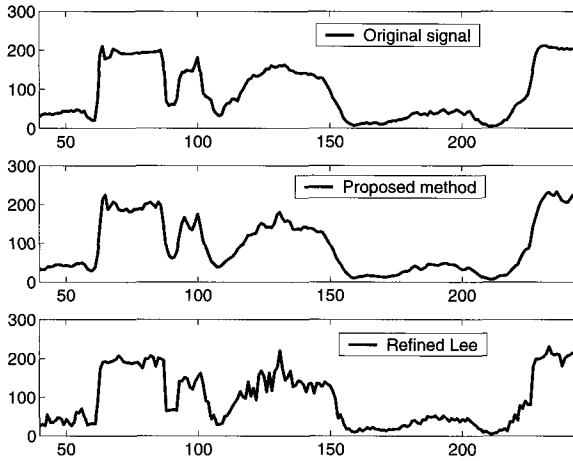


Fig. 9. Profile comparison (column 185) for the Lena image. Figure is from [53], ©2002 IEEE, used with permission.

### Real SAR images

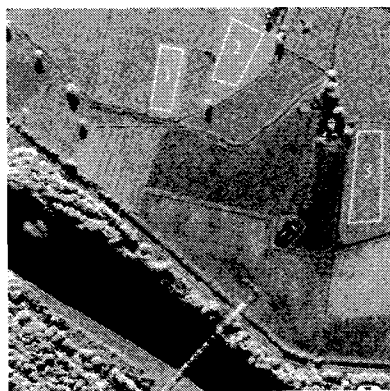
Shown in Fig. 10(a) is a  $400 \times 400$  Ku-band SAR image over the Rio Grande river near Albuquerque, New Mexico, acquired by the Sandia National Laboratories Twin Otter aircraft [43]. An enlarged part of the despeckled image is shown in Fig. 10(b)-(e) for the three wavelet-based filters and the refined Lee filter, respectively. The wavelet transform was performed with four levels of decomposition and the cycle spinning technique was applied with  $4 \times 4$  shifts.

Since the noise-free image is not available, we only use the ENL to assess the filter's noise reduction performance in homogeneous areas. As illustrated in Fig. 10(a), three uniform areas are selected for the quantitative analysis. Table 7 lists the ENL values before and after filtering. In all cases, the proposed filter outperforms the other three filters with the highest ENL values, which is consistent with the simulation results. The proposed method provides comparable edge preservation performance as the other two wavelet-based methods, however the refined Lee filter is able to reconstruct edges with better visual quality.

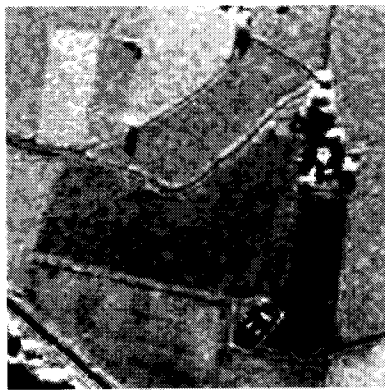
### 4. Speckle Reduction Using a Low-Complexity Wavelet Denoising Process

While considerable effort has been devoted to adopting more and more complicated wavelet denoising algorithms for speckle reduction, in which the logarithmic transform is applied, so far there exist only a few studies where the problem of wavelet despeckling without performing the logarithmic transform has been addressed and investigated. In [11], the authors showed how a signal corrupted by multiplicative noise can be expressed mathematically as a signal with additive noise, which suggests that with such an approach, wavelet-based filters originally designed for additive noise should perform equally well for SAR speckle reduction. The authors then propose a wavelet-based filter that works directly on SAR images containing multiplicative speckle noise. The filter differs from most existing wavelet-

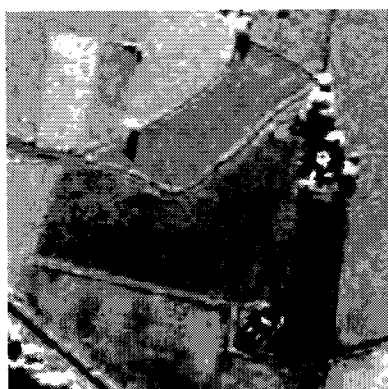




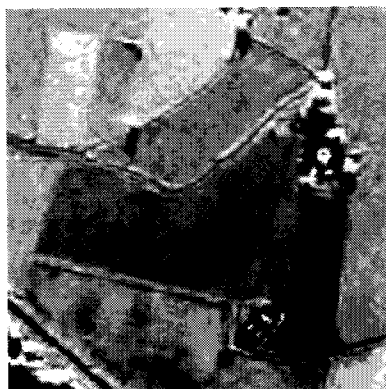
(a) SAR image



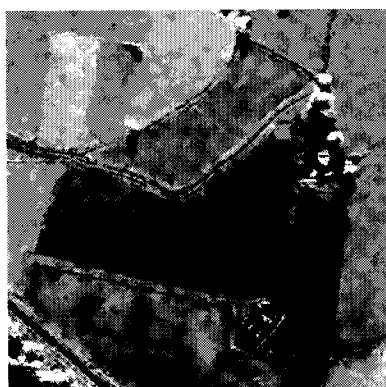
(b) Bayesian soft thresholding



(c) Bayesian MMSE estimation



(d) The proposed method



(e) Refined Lee

Fig. 10. Comparison of different despeckling methods for the SAR image. Figure is from [53], ©2002 IEEE, used with permission.

based filters in that noise reduction is achieved via a recursive use of the wavelet transform smoothing function. Another example of a wavelet-based speckle filter that does not involve a log transformation is presented in [18], where the possibility of feature preservation by selectively thresholding the wavelet coefficients is investigated. After the wavelet transform, the algorithm classifies each pixel in each detailed image as an “edge” or “non-edge” pixel based on a predefined threshold. Wavelet denoising is then accomplished by applying different shrinkage factors to “edge” and “non-edge” pixels. Algorithms in [11] and [18] are both highly heuristic in nature. Although experimental results show that they both provide satisfactory despeckling performance, the limitation arises from the fact that neither one takes advantage of the known statistical properties of speckle noise. For example, in [11] the degree of feature preservation is controlled mainly by a threshold applied at the wavelet transform level. Wavelet despeckling without a log-transform can also be seen in very recent work of Foucher, *et al.* [16]. The authors assume the Gamma distribution for both the speckle and the underlying radar reflectivity, and then use the Pearson distribution to approximate the probability density function (pdf) of the wavelet coefficients. Among wavelet-based despeckling approaches that do not involve a log transformation, the technique presented in [16] is the first attempt we are aware of where MAP-based wavelet despeckling uses a realistic prior model.

Because of the aforementioned problems associated with the described methods, an efficient and yet rigorous approach based on the multiplicative model is still lacking. In [39], a simple wavelet denoising algorithm is developed for images corrupted by film-grain type noise which is additive but signal-dependent. In this section, we will extend that method to SAR images and develop a low-complexity speckle reduction method without performing a log-transformation and the resultant mean adjustment and exponential operations. The difference between the proposed algorithm and the one in [39] are: 1) we focus on the development of a filter for SAR images with the inherent multiplicative speckle noise model, whereas in [39] the images under analysis had an additive observation model with signal-dependent noise; and 2) we estimate the shrinkage factor from the support region of the wavelet basis functions rather than the point-wise approximation used in [39]. Experimental results demonstrate that the denoising performance can be improved significantly by these modifications.

In what follows, first, we derive the shrinkage factor based on the MMSE (Minimum Mean Square Error) estimator in the wavelet domain for a general additive noise model. Next, we introduce the wavelet denoising algorithm developed in [39] for film-grain noise, then we modify this algorithm to SAR speckle reduction. Lastly, we quantify the improvement of the proposed approach over the method presented in [39]. In addition, we compare it with a standard spatial speckle filter, as well as another log-transform-based wavelet despeckling algorithm that has comparable computational complexity.

In the context of wavelet denoising (see equations 1 and 2), retrieving  $\mathbf{f}$  from the noisy observation  $\mathbf{w}$  is generally accomplished by shrinking the wavelet coefficients  $\mathbf{y}$ . If we apply the MMSE technique to wavelet denoising, our goal is to seek optimal shrinkage factors  $\{\eta_{i,j}\}$  that minimize the mean square error between the estimated and original wavelet

coefficients. That is:

$$\hat{x}_{i,j} = \eta_{i,j} y_{i,j}, \quad (32)$$

$$\eta_{i,j} = \arg \min_{\eta_{i,j}} E\{(x_{i,j} - \hat{x}_{i,j})^2\}. \quad (33)$$

The optimal shrinkage factor  $\eta_{i,j}$  can be sought by solving:

$$\frac{\partial}{\partial \eta_{i,j}} E\{(x_{i,j} - \eta_{i,j} y_{i,j})^2\} = 0. \quad (34)$$

By expanding the above formula, the MMSE solution of  $\eta_{i,j}$  can be derived as:

$$\eta_{i,j} = \frac{E(x_{i,j} y_{i,j})}{E(y_{i,j}^2)} = \frac{E(y_{i,j}^2) - E(n_{i,j} y_{i,j})}{E(y_{i,j}^2)}. \quad (35)$$

For wavelet coefficients representing homogeneous areas, we expect the value of  $\eta$  to be close to zero to reduce noise. On the other hand, for edges or real textures, the value of  $\eta$  should approach one to preserve image details.

#### 4.1. Formulation

Film-grain noise usually occurs when scanning an image recorded on photographic film. An observed image corrupted by this type of noise is modeled by:

$$\mathbf{w} = \mathbf{f} + f^\gamma \epsilon, \quad (36)$$

where  $\epsilon$  is AWGN with zero mean and variance  $\sigma_n^2$ ,  $\mathbf{f}$  is the original signal, and  $\gamma$  is a constant between 1/3 and 1/2. The authors of [39] assumed the wavelet basis functions to be short enough so that  $\mathbf{f}$  can be treated as constant within the support region of the basis functions. The optimal shrinkage factor was derived as:

$$\eta_{i,j} = \frac{E(y_{i,j}^2) - \sigma_n^2 x_{i,j}^{2\gamma}}{E(y_{i,j}^2)}, \quad (37)$$

where  $E(y_{i,j}^2)$  is estimated by:

$$E(y_{i,j}^2) \approx y_{i,j}^2. \quad (38)$$

For SAR imagery, we have a multiplicative speckle model:

$$\mathbf{w} = \mathbf{f} \epsilon, \quad (39)$$

where  $\mathbf{f}$  is the noise-free signal,  $\mathbf{w}$  is the observation and  $\epsilon$  is the speckle random noise with unit mean. SAR images are usually produced in two formats. In intensity format, the random variable  $\epsilon$  follows the Gamma distribution, whereas in amplitude format, it obeys a multi-convolution of the Rayleigh probability density function [48]. We can easily decompose the multiplicative noise model in (39) into an additive model by expressing it in the following form:

$$\mathbf{w} = \mathbf{f} + \mathbf{f} (\epsilon - 1), \quad (40)$$

with the DWT of  $\mathbf{y} = \mathbf{x} + \mathbf{N}$ , and where  $(\epsilon - 1)$  is a random variable with zero mean, and  $\mathbf{f}(\epsilon - 1)$  is the signal-dependent noise imposed on the underlying unknown signal  $f$ .

In [39], except for zero mean and being independent of the original signal, no other statistical property of the noise  $\epsilon$  in (36) plays a role in the derivation of the shrinkage factor given by (37). As a result, if  $\mathbf{N}$  denotes the wavelet transform of the signal-dependent noise  $\mathbf{f}(\epsilon - 1)$ , the optimal shrinkage factor for SAR images can be deduced in a similar manner as [39]. The core of Öktem’s method is represented by (37), in which if we set  $\gamma = 1$ , and replace  $\sigma_n$  with  $\sigma_f$  (the speckle noise standard deviation), then that method can be directly extended to SAR speckle reduction. The primary advantage of Öktem’s method is its simplicity due to the fact that both  $E(y_{i,j}^2)$  and  $E(N_{i,j}y_{i,j})$  are estimated in a point-wise fashion. In what follows, we will expand their method by incorporating spatial information into the estimation process and apply it to the problem of SAR speckle reduction.

For the wavelet coefficient at pixel  $(i, j)$ , we use neighboring pixels to estimate its variance

$$E(y_{i,j}^2) = \frac{1}{D_{\mathcal{N}}} \sum_{(p,q) \in \mathcal{N}(i,j)} y_{p,q}^2, \tag{41}$$

where  $\mathcal{N}(i, j)$  denotes the neighborhood system of pixel  $(i, j)$ , and  $D_{\mathcal{N}}$  is the corresponding dimension.

At level 1 of a 2-D dyadic wavelet transform, we have:

$$N_{1,(\overline{i},\overline{j})} \text{DWT}\{\mathbf{f}(\epsilon - 1)\}_{i,j} = \sum_{p=0}^{M-1} \sum_{q=0}^{M-1} \tilde{h}_p \tilde{h}_q f_{2i+p,2j+q} (\epsilon_{2i+p,2j+q} - 1), \tag{42}$$

$$y_{1,(\overline{i},\overline{j})} \text{DWT}\{\mathbf{f} \epsilon\}_{i,j} = \sum_{p=0}^{M-1} \sum_{q=0}^{M-1} \tilde{h}_p \tilde{h}_q f_{2i+p,2j+q} \epsilon_{2i+p,2j+q}, \tag{43}$$

where  $\tilde{h}$  ( $\tilde{h} = h$  or  $g$ ) and  $\tilde{\tilde{h}}$  ( $\tilde{\tilde{h}} = h$  or  $g$ ) denote the two wavelet filters applied to the row and column directions, respectively, and  $M$  is the length of the wavelet filters. Using the above two formulas, the correlation function between  $N_{1,(i,j)}$  and  $y_{1,(i,j)}$  can be expressed by:

$$E(N_{1,(i,j)}y_{1,(i,\overline{j})}) \sum_{p=0}^{M-1} \sum_{q=0}^{M-1} \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} \tilde{h}_p \tilde{h}_q \tilde{\tilde{h}}_m \tilde{\tilde{h}}_n \cdot E(f_{2i+p,2j+q}(\epsilon_{2i+p,2j+q} - 1)f_{2i+m,2j+n}\epsilon_{2i+m,2j+n}). \tag{44}$$

Since speckle  $\epsilon$  is independently distributed with unit mean, we obtain:

$$\begin{aligned} E(N_{1,(i,j)}y_{1,(i,j)}) &= \sum_{p=0}^{M-1} \sum_{q=0}^{M-1} \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} \tilde{h}_p \tilde{h}_q \tilde{\tilde{h}}_m \tilde{\tilde{h}}_n f_{2i+p,2j+q} f_{2i+m,2j+n} \\ &\quad \cdot E(\epsilon^2 - \epsilon)\delta(p - m)\delta(q - n) \\ &= \sigma_f^2 \sum_{p=0}^{M-1} \sum_{q=0}^{M-1} \tilde{h}_p^2 \tilde{h}_q^2 f_{2i+p,2j+q}^2, \end{aligned} \tag{45}$$

where  $\delta(\cdot)$  is the discrete delta function, and  $\sigma_f^2$  is the variance of the speckle noise. For SAR images in the intensity format

$t$ ,  $\sigma_f^2 = 1/L$ ; while for the amplitude format,  $\sigma_f^2 = (\frac{4}{\pi} - 1)/L$  [48]. In both cases,  $L$  is the number of looks.

We assume that the wavelet basis functions are short enough so that  $\mathbf{f}$  can be treated as constant within the support region of the basis functions. By using the orthogonal property of the wavelet filter:

$$\sum_k^{M-1} g_k g_{k+2l} = \delta(l), \quad \sum_k^{M-1} h_k h_{k+2l} = \delta(l), \quad (46)$$

we can prove that at level  $l > 1$ ,

$$E(N_{l,(i,j)} y_{l,(i,j)}) \approx \sigma_f^2 \sum_{p=0}^{M-1} \sum_{q=0}^{M-1} \tilde{h}_p^2 \tilde{h}_q^2 f_{2^l i + 2^{l-1} p, 2^l j + 2^{l-1} q}^2, \quad l > 1 \quad (47)$$

See our paper [55] for the details.

As seen both in (37) and (47), the unknown noise-free signal  $\mathbf{f}$  is a prerequisite for computing the shrinkage factor for every wavelet coefficient. We apply the shrinkage function only to wavelet subbands  $HH$ ,  $HL$  and  $LH$ , in the meantime we incorporate the denoised approximation subband  $\widehat{LL}$  into the denoising procedure to approximate the noise-free signal  $\mathbf{f}$ .

Due to the admissibility condition that the scaling filter  $g$  must satisfy, we estimate the original signal at level  $l$  by linearly decreasing  $\widehat{LL}_l$  by a factor of  $(\sqrt{2})^{2l}$ :

$$\hat{f}_{2^l i + 2^{l-1} p, 2^l j + 2^{l-1} q} \approx \frac{1}{(\sqrt{2})^{2l}} \widehat{LL}_{l,(i+\lfloor \frac{p}{2} \rfloor, j+\lfloor \frac{q}{2} \rfloor)}, \quad l = 1, 2, \dots, \mathcal{L}, \quad (48)$$

where the  $\lfloor \cdot \rfloor$  function rounds the argument to the nearest integer towards minus infinity. The power of two is to take account of the admissibility condition in both the row and column directions.

The overall image despeckling procedure is summarized as follows:

- (1) Perform a  $\mathcal{L}$ -level 2-D discrete wavelet transform.
- (2) Start from level  $\mathcal{L}$ . At level  $l$ , for each wavelet coefficient  $y_{l,(i,j)}$  in each wavelet subband of  $\{LH_l, HL_l, HH_l\}$ :
  - Estimate the local variance  $E(Y_{l,(i,j)}^2)$  using (41);
  - Approximate the original signal  $\hat{\mathbf{f}}$  at the current level by the denoised approximation subband  $\widehat{LL}_l$  using (48) (at level  $\mathcal{L}$ , i.e, the coarsest resolution,  $\widehat{LL}_{\mathcal{L}} = LL_{\mathcal{L}}$ ). Then estimate the correlation function  $E(N_{l,(i,j)} y_{l,(i,j)})$  using (45) or (47);
  - Calculate the shrinkage factor  $\eta_{l,(i,j)}$  according to (35) for each individual wavelet coefficient;
  - Shrink the noisy wavelet coefficient  $y_{l,(i,j)}$  by:

$$\hat{x}_{l,(i,j)} = \max(0, \eta_{l,(i,j)}) y_{l,(i,j)}, \quad \hat{x}_l = \{\widehat{HL}_l, \widehat{HL}_l, \widehat{HH}_l\}. \quad (49)$$

(3) Perform a 2-D inverse wavelet transform at the current level:

$$\widehat{LL}_{l-1} = \text{IDWT}(\widehat{LL}_l, \widehat{LH}_l, \widehat{HL}_l, \widehat{HH}_l). \quad (50)$$

The resulting approximation subband  $\widehat{LL}_{l-1}$  will be used at the next level.

(4) Repeat step 2 and step 3 until level 0 is reached.

The above procedure clearly demonstrates the difference between this algorithm and the one proposed in the previous section. For the latter, denoising each detail subband requires the denoised approximation subband at the current level that is a result from the immediate upper level within the DWT framework, therefore it can only be performed in a sequential manner. On the contrary, the former algorithm can be naturally structured in a parallel way since denoising any specific detail subband needs no information from other subbands.

#### 4.2. Experimental Results and Discussions

In this section, we illustrate the performance of our proposed wavelet despeckling algorithm for simulated SAR images in both intensity and amplitude formats.

As shown in Fig. 11, four standard test images, including *House*, *Peppers*, *Lena* and *Barbara* ordered with increasing degree of texture content, were used for comparison. All images have dimensions of  $256 \times 256$ . With regard to the wavelet family, the length of the wavelet filter, the decomposition level and the image border extension mode, we employed exactly the same settings as used in the previous section. In addition to the method developed by Öktem, we also compare our method with two other despeckling methods described briefly below.

In [37], a low-complexity wavelet denoising algorithm called LAWML (Locally Adaptive Window based denoising using ML), is developed for reducing additive white Gaussian noise. In the algorithm, wavelet coefficients  $y$  are modeled as conditionally independent zero-mean Gaussian random variables given their variances. For the wavelet coefficient at pixel  $(i, j)$  denoted by  $y_{i,j}$ , the variance is estimated in a local neighborhood:

$$\sigma_{i,j}^2 = \max \left( 0, \frac{1}{D_{\mathcal{N}}} \sum_{(p,q) \in \mathcal{N}(i,j)} y_{p,q}^2 - \sigma_n^2 \right), \quad (51)$$

where  $\mathcal{N}(i, j)$  and  $D_{\mathcal{N}}$  are defined as in (41) and  $\sigma_n^2$  is the variance of the noise. Then  $\sigma_{i,j}^2$  is used in the MMSE estimator to obtain the noise-free wavelet coefficients  $\hat{x}_{i,j}$ :

$$\hat{x}_{i,j} = \frac{\sigma_{i,j}^2}{\sigma_{i,j}^2 + \sigma_n^2} y_{i,j}. \quad (52)$$

As (35) and (52) both resemble the classic Wiener filter in the wavelet domain, we form another despeckling method using LAWML as a kernel. This despeckling method starts with the conventional logarithmic transform, and is concluded by the mean bias correction and the exponential operation.

To make the connection with those statistical filters that are applied directly in the image domain, we also include the refined Lee filter as a reference against which to assess the performance of the proposed speckle filter.

We corrupted the test images by multiplying simulated spatially uncorrelated speckle noise. Quantitative comparison is carried out in the sense of the signal-to-noise ratio (SNR).

For simplicity, we only deploy  $3 \times 3$ ,  $5 \times 5$  and  $7 \times 7$  local square windows to estimate the local variance of the wavelet coefficient  $E(y_{i,j}^2)$ . Our results indicate that a local window with size  $5 \times 5$  offers the best performance in terms of the tradeoff between noise suppression and detail preservation. We therefore only report the results using the  $5 \times 5$  window. The SNR results of applying these four despeckling algorithms to two of the four test images are listed in Tables 1.8 and 1.9. Fig. 11 presents the visual comparison of the four speckled test images with speckle noise ( $L = 3$ ) in amplitude format and their respective denoised versions by the proposed method, the LAWML method and the refined Lee filters, respectively. Based on these results, we note the following observations:

- Under all conditions, the proposed method achieves remarkable improvement over the method by Öktem *et al.*
- For amplitude SAR images, the proposed method outperforms both the refined Lee filter and the LAWML.
- For single-look intensity SAR images, the refined Lee filter provides better performance than any of the three wavelet-based filters.
- The proposed method is based on the assumption that the underlying signal is piecewise constant with respect to the support region of the wavelet filters. This is not true for images containing high frequency textural variations such as the *Barbara* image, which explains why the proposed method only achieves comparable performance to the LAWML method for intensity images. On the other hand, for images mainly containing homogeneous areas (*House* and *Peppers*) and images with medium texture contents (*Lena*), for which our assumption is valid, the proposed method achieves much better performance than the other two methods.
- For the *Barbara* image, some small-scale structure has been lost in the reconstruction by the refined Lee filter, as expected from its less spatially adaptive template over which the weighted averaging is performed. On the other hand, the wavelet multiscale denoising algorithms appear to give better structure preservation in the area of excess variations.

As mentioned in the previous section, the discrete wavelet transform is a shift-variant system due to the downsampling operation. We have applied a cycle spinning technique [7], which reduces artifacts like specks and Gibbs phenomena, and consequently the denoising performance can be further improved. We have used a total of  $4 \times 4$  shifts in both rows and columns, as before.

The proposed algorithm requires more computational time at each wavelet coefficient than LAWML when computing the shrinkage factor. If the wavelet filter has a length  $N$ , the proposed method needs  $N^2$  more multiplications and  $(N^2 - 1)$  more additions for

each wavelet coefficient. All the computation was carried out using programs written in C and running on a SUN Ultra 1 machine. In the framework of a three-level DWT with Daubechies' filter of length 6, the CPU time for despeckling the  $256 \times 256$  *Lena* image is about 1.1s, 2.4s, and 4.0s for the method developed by Öktem, the method based on LAWML, and the proposed method, respectively. The refined Lee filter requires about 2.0s.

Looks	1	2	3	16	25	36
	Intensity format					
This Study	4.36	6.81	<b>8.23</b>	<b>13.37</b>	<b>14.73</b>	<b>15.74</b>
The method based on [39]	0.72	3.97	5.65	11.79	13.26	14.46
The method based on LAWML	2.93	6.23	7.63	12.85	14.16	15.23
Refined Lee	<b>4.99</b>	<b>6.85</b>	7.95	12.76	14.04	15.08
	Amplitude format					
This Study	<b>8.87</b>	<b>11.01</b>	<b>12.20</b>	<b>17.29</b>	<b>18.57</b>	<b>19.75</b>
The method based on [39]	6.52	9.13	10.50	16.11	17.65	18.93
The method based on LAWML	7.52	10.20	11.49	16.66	18.04	19.18
Refined Lee	8.56	10.50	11.63	16.52	17.84	18.93

Looks	1	2	3	16	25	36
	Intensity format					
This study	2.07	<b>4.22</b>	5.35	<b>9.51</b>	<b>10.64</b>	<b>11.55</b>
The method based on [39]	-0.12	2.73	4.18	9.00	10.21	11.19
The method based on LAWML	1.34	4.17	<b>5.38</b>	9.49	10.58	11.52
Refined Lee	<b>2.56</b>	3.99	4.85	8.65	9.80	10.79
	Amplitude format					
This Study	<b>5.86</b>	<b>7.63</b>	<b>8.60</b>	<b>12.86</b>	<b>14.12</b>	<b>15.22</b>
The method based on [39]	4.93	6.89	7.99	12.54	13.86	15.05
The method based on LAWML	5.29	7.42	8.49	12.76	14.03	15.11
Refined Lee	5.35	6.83	7.74	12.18	13.54	14.73

## References

1. E. H. L. Aarts, *Simulated annealing and Boltzmann machines: a stochastic approach to combinatorial optimization and neural computing*, New York: Wiley, 1989.
2. J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society, Series B*, vol. 48, no. 3, pp. 259–302, 1986.
3. S. G. Chang and M. Vetterli, "Spatial adaptive wavelet thresholding for image denoising," in *Proceedings of IEEE International Symposium On Image Processing*, volume 1, pp. 374–377, 1997.



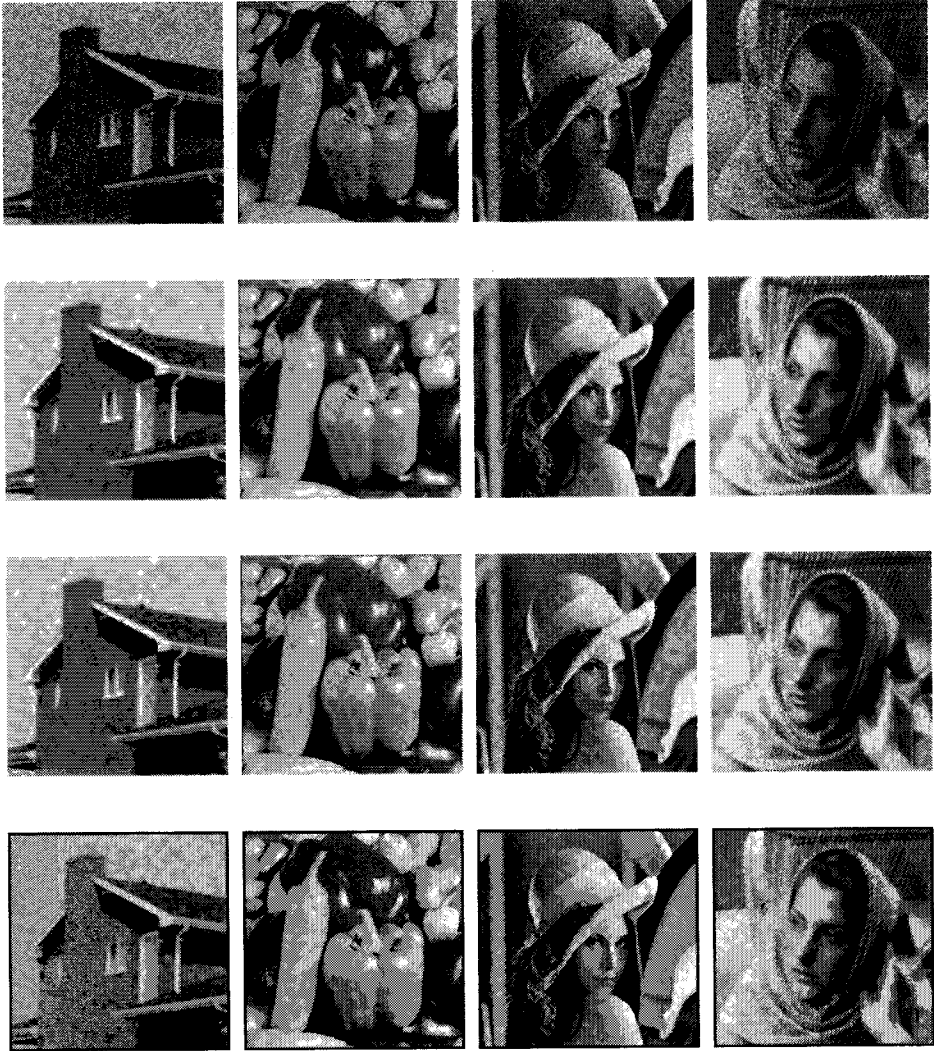


Fig. 11. From left to right: *House* image, *Peppers* image, *Lena* image, and *Barbara* image. From top to bottom: Speckled images with 3-look speckle in amplitude format, despeckled images by the proposed method, by the method based on LAWML, and by the refined Lee filter. Figure is from [55], ©2003 IEEE, used with permission.

4. S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transaction on Image Processing*, vol. 9, no. 9, pp. 1532–1546, September 2000.
5. S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Transaction on Image Processing*, vol. 9, no. 9, pp. 1522–1531, September 2000.
6. H. Chipman, E. Kolaczyk, and R. McCulloch, "Adaptive Bayesian wavelet shrinkage," *Journal*

- of *American Statist. Assoc.*, vol. 92, pp. 1413–1421, 1997.
7. R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics*, A. Antoniadis and G. Oppengeim, editors, pp. 125–150, Springer-Verlag, Berlin, Germany, 1995.
  8. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE trans. on Signal Processing, Special issue on Wavelets and Filterbanks*, vol. 46, no. 4, pp. 886–902, April 1998.
  9. I. Daubechies, "Ten lectures on wavelets," *SIAM*, 1992.
  10. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 29, no. 1, pp. 1–28, 1977.
  11. Y. Dong, B. C. Forster, A. K. Milne, and G. A. Morgan, "Speckle suppression using recursive wavelet transforms," *International Journal of Remote Sensing*, vol. 19, no. 2, pp. 317–330, 1998.
  12. Y. Dong, A. K. Milne, and B. C. Forster, "Toward edge sharpening: a SAR speckle filtering algorithm," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 39, no. 4, pp. 851–863, April 2001.
  13. D. L. Donoho, "De-noising by soft thresholding," *IEEE Trans. on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
  14. D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
  15. R. C. Dubes and A. K. Jain, "Random field models in image analysis," *J. Appl. Statist.*, vol. 16, pp. 131–163, 1989.
  16. S. Foucher, G. B. Benie, and J. Boucher, "Multiscale MAP filtering of SAR images," *IEEE Transaction on Image Processing*, vol. 10, no. 1, pp. 49–60, January 2001.
  17. V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 4, no. 2, pp. 157–166, March 1982.
  18. S. Fukuda and H. Hirotsawa, "Suppression of speckle in synthetic aperture radar images using wavelet," *International Journal of Remote Sensing*, vol. 19, no. 3, pp. 507–519, 1998.
  19. L. Gagnon and A. Jouan, "Speckle filtering of SAR images—a comparative study between complex-wavelet-based and standard filters," in *Proceedings of SPIE*, volume 3169, pp. 80–91, 1997.
  20. J. W. Goodman, *Statistical Properties of Laser Speckle Patterns*, Springer-Verlag, Heidelberg, Germany, 1980.
  21. H. Guo, J. E. Odegard, M. Lang, R. A. Gopinath, I. W. Seleshick, and C. S. Burrus, "Wavelet based speckle reduction with application to SAR based ATD/R," in *Proceedings of the 1994 IEEE International Conference on Image Processing*, volume 1, pp. 75–79, 1994.
  22. W. Hagg and M. Sties, "Efficient speckle filtering of SAR images," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, volume 3, pp. 2140–2142, Pasadena, CA, USA, 1994.
  23. E. Jakeman and P. N. Pusey, "A model for non-Rayleigh sea echo," *IEEE Transaction on Antennas Propagation*, vol. 24, pp. 806–814, 1976.
  24. Z. Kato, *Multi-scale Markovian modelization in computer vision with applications to SPOT image segmentation*, PhD thesis, University of Nice, France, December 1994.
  25. D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive restoration of images with speckle," *IEEE Transaction on Accoustic, Speech and Signal Processing*, vol. ASSP-35, no. 3, pp. 373–382, March 1987.
  26. J. S. Lee, "Refined filtering of image noise using local statistics," *Computer Graphics Image Processing*, vol. 15, no. 14, pp. 380–389, September 1981.
  27. J. S. Lee, "Speckle suppression and analysis for synthetic aperture radar images," *Optical Engi-*

- neering, vol. 25, no. 5, pp. 636–643, May 1986.
28. J. S. Lee, M. R. Grunes, and S. A. Mango, "Speckle reduction in multipolarization, multifrequency SAR imagery," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 29, no. 4, pp. 535–544, July 1991.
  29. C. Li, "Two adaptive filters for speckle reduction in SAR images by using the variance ratio," *Int. J. Remote Sensing*, vol. 9, no. 4, pp. 641–653, 1988.
  30. S. Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, Tokyo, 1995.
  31. Q. Lin and J. P. Allebach, "Combating speckle in SAR images: Vector filtering and sequential classification based on a multiplicative noise model," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 28, no. 4, pp. 647–653, July 1990.
  32. J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Transaction on Image Processing*, vol. 10, no. 11, pp. 1647–1658, November 2001.
  33. A. Lopes, R. Touzi, and E. Nezry, "Adaptive speckle filters and scene heterogeneity," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 28, no. 6, pp. 992–1000, November 1990.
  34. M. Malfait and D. Roose, "Wavelet-based image denoising using a Markov random field *a priori* model," *IEEE Transaction on Image Processing*, vol. 6, no. 4, pp. 549–565, April 1997.
  35. D. Marr, *Vision*, Freeman Press, San Fransisco, CA, 1982.
  36. M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. on Patt. Anal. and Mach. Intel.*, vol. 11, no. 7, pp. 674–693, July 1989.
  37. M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Trans. on Signal Processing Letters*, vol. 6, no. 12, pp. 300–303, December 1999.
  38. L. M. Novak and M. C. Burl, "Optimal speckle reduction in polarimetric SAR imagery," *IEEE Transaction on Aerosp. Electron. Syst.*, vol. 26, no. 2, pp. 293–305, February 1990.
  39. R. Öktem and K. Egiazarian, "A transform domain algorithm for filtering film-grain type noise," in *Proceedings of International Conference on Information, Communications and Signal Processing*, pp. 75–79, Singapore, December 1999.
  40. C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, Artech House, Boston, 1998.
  41. J. Park, W. Song, and W. Pearlman, "Speckle filtering of SAR images based on adaptive windowing," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 146, no. 4, pp. 191–197, August 1999.
  42. R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol. 26, no. 2, pp. 195–239, 1984.
  43. <http://www.sandia.gov/radar/imageryku.html>, SAR Programs, Sandia National Laboratories.
  44. E. P. Simoncelli, "Bayesian denoising of visual images in the wavelet domain," in *Bayesian Inference in Wavelet Based Models*, pp. 291–308, Springer-Verlag, New York, 1999.
  45. A. H. S. Solberg, T. Taxt, and A. K. Jian, "A Markov random field model for classification of multisource satellite imagery," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 34, no. 1, pp. 100–112, January 1996.
  46. J. R. Sveinsson and J. A. Benediktsson, "Speckle reduction of SAR images using wavelet-domain hidden Markov models," in *Proceedings of IEEE Intenational Geoscience and Remote Sensing Symposium*, volume 4, pp. 1666–1668, Honolulu, HI, 2000.
  47. R. Touzi, "A protocol for speckle filtering of SAR images," in *Proceedings of CEOS Workshop*, Toulouse, France, October 1999.
  48. F. Ulaby and M. C. Dobson, *Handbook of Radar Scattering Statistics for Terrain*, Artech House Publishing Co., Northwood, Mass., 1989.
  49. H. Wakabayashi and K. Arai, "Method of speckle noise reduction for SAR data," *International*

*Journal of Remote Sensing*, vol. 17, no. 10, pp. 1837–1849, July 1996.

50. M. Walesa and M. Datcu, "Model-based despeckling and information extraction from SAR images," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 38, no. 5, pp. 2258–2268, September 2000.
51. R. G. White, "A simulated annealing algorithm for radar cross-section estimation and segmentation," in *SPIE Int. Conf on Applications of Artificial Neural Networks V*, volume 2243, pp. 231–241, Orlando, FL, 1994.
52. Y. Wu and H. Maitre, "Smoothing speckled synthetic aperture radar images by using maximum homogeneous region filters," *Optical Engineering*, vol. 31, no. 8, pp. 1785–1792, August 1992.
53. H. Xie, L. E. Pierce, and F. T. Ulaby, "Speckle reduction using wavelet denoising and Markov random field modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, , September 2002.
54. H. Xie, L. E. Pierce, and F. T. Ulaby, "Statistical properties of logarithmically transformed speckle," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 3, pp. 721–727, March 2002.
55. H. Xie, F. T. Ulaby, and L. E. Pierce, "SAR image speckle reduction using a low-complexity wavelet denoising process," *submitted to IEEE Transactions on Geoscience and Remote Sensing*, April 2002.
56. X. Zong, A. F. Laine, and E. A. Geiser, "Speckle reduction and contrast enhancement of echocardiograms via multiscale nonlinear processing," *IEEE Transaction on Medical Imaging*, vol. 17, no. 4, pp. 532–540, August 1998.

## CHAPTER 9

# A WAVELET REPRESENTATION OF MULTISPECTRAL IMAGES

Paul Scheunders

*Vision Lab, Department of Physics, University of Antwerp  
Groenenborgerlaan 171, 2020 Antwerpen, Belgium  
E-mail: scheun@ruca.ua.ac.be*

In this work, a wavelet representation of multispectral images is presented. The representation is based on a multiresolution extension of the First Fundamental Form that accesses gradient information of vector-valued images. With the extension, multiscale edge information of multispectral images is extracted. Moreover, a wavelet representation is obtained that, after inverse transformation, accumulates all edge information in a single greylevel image. In this work, a redundant wavelet representation is presented using dyadic wavelet frames. It is then extended towards orthogonal wavelet bases using the Discrete Wavelet Transformation (DWT). The representation is shown to be a natural framework for image fusion. An algorithm is presented for fusion and merging of multispectral images. The concept is successfully applied to the problem of multispectral and hyperspectral image merging.

### 1. Introduction

With the evolution of imaging technology, an increasing number of imaging modalities becomes available. In remote sensing, sensors are available that measure light reflectance at different wavelengths simultaneously. Multispectral images are generated by positioning each measurement in a separate band. When the number of bands is so high that a more or less continuous reflectance spectrum is obtained, one refers to hyperspectral imagery.

Much work has been devoted to the classification and/or segmentation of multispectral images for identification purposes.<sup>1,2</sup> However, usually only a fraction of the data provides unique or useful information. Moreover, reducing the data set reduces the complexity of a classification procedure. For

this reason, it is common practice to select a limited number of bands or extract a limited number of features from the entire multispectral set. Also a fusion of several bands into one or a few greylevel images is performed.<sup>3,4</sup> A similar procedure merges different images (e.g. a high resolution panchromatic image with a low resolution multispectral image) for enhancement.<sup>5</sup> Other image processing procedures to facilitate the classification and segmentation of multispectral images include noise filtering<sup>6,7,8</sup> and image enhancement.<sup>9,10,11</sup>

It is obvious that all these image processing and analysis techniques would benefit from the combined processing of the different bands involved. Nevertheless, in most cases single-valued processing and analysis techniques are applied to each of the bands separately. The results for each band are then combined in a usually heuristic manner.

A large part of image processing and analysis techniques makes use of the image edge information, that is contained in the image gradient. A nice way of describing multispectral edges was given<sup>12</sup>. Here, the images "first fundamental form", a quadratic form, is defined for each image point. This is a local measure of directional contrast based upon the gradients of the image bands. This measure is maximal, at each image point, in a particular direction, that in the greylevel case is the direction of the gradient. Based on this definition, a color edge detection algorithm was described<sup>13</sup> and a color image anisotropic diffusion algorithm was described<sup>14</sup>.

In single-valued images, multiresolution techniques are used to describe edges. The wavelet transform e.g. is successfully applied to compression, noise-reduction, enhancement, classification and segmentation of greylevel images. However, when applied to multispectral images, it is applied to each band separately.

In this paper, a new multispectral image wavelet representation is presented. This representation allows for a multiscale edge description of multispectral images. The idea for the representation is based on the first fundamental form<sup>12</sup> and the dyadic wavelet representation of Mallat.<sup>15</sup> The latter decomposes an image into detail images that are convolutions of the image with derivatives of a smoothing function. These detail images can be written as the derivatives of the image, smoothed at different scales. This observation allows for a definition of multiscale fundamental forms. The eigenvectors and eigenvalues of these quadratic forms describe the directions and rates of change of the multispectral image at that particular scale.

This extension has the advantage of allowing for a reconstruction to-

wards a single-valued image, combining multiscale edge information of all bands involved<sup>16</sup>. Applications were developed for multispectral image fusion and merging<sup>17</sup>, anisotropic diffusion filtering of color images<sup>18</sup>, color image segmentation<sup>19</sup> and image denoising.<sup>20</sup>

A disadvantage of the presented technique is the use of a redundant wavelet transform with large computing power and memory requirements. For this reason, an orthogonal multispectral wavelet representation will be developed as well, based on the Discrete Wavelet Transform (DWT). For this, the concept of maximal gradient is generalized towards linear vector operators in the image plane with equal components along rows and columns. The generally applied separable conjugate mirror filter decomposition of the DWT is modified in order to include such operators. In this way, an orthogonal wavelet representation of multispectral images is obtained such that detail images from different bands are combined into one single representation.

To demonstrate the proposed representations, we will apply the concepts to the problem of fusion and merging of multispectral images. A recent overview of the problem of multispectral image fusion and merging is given<sup>21</sup>. We define image fusion as the combination of several bands of a multispectral image into one greylevel image. Applications are image enhancement for visualization and reduction of the complexity of classification tasks.<sup>22,23,24,25</sup> Another important application in the literature is the fusion of multisensor imagery as for instance provided by ground based or airborne (military) platforms and surveillance devices.<sup>26,27,25</sup>

We will refer to image merging as the process of combining a greylevel image with each band of a multispectral image in order to improve the spatial resolution of the multispectral image. Applications are the combination of a high-resolution greylevel image with a low-resolution multispectral image to obtain high-resolution multispectral information.<sup>5,28,29</sup> An example is given by the merging of SPOT Panchromatic data with Landsat Thematic Mapper multispectral images.<sup>30,31,32,33,34</sup>

Most of the fusion and merging techniques described in the literature are pixel-based. Many techniques are based on multiresolution processing. The multiresolution approach allows for a combination of edge information at different scales. A very popular paradigm is given by the wavelet transform.<sup>5,22,31,25</sup> Other methods, like pyramid-based fusion were also described.<sup>35,36</sup> The rule for combining the detail information is an important issue. The most common rule for fusion is to take the detail coefficient from one of the bands (e.g. the one with highest energy). For merging the

most common rule is substitution (e.g. substitution of the detail images of a high-resolution greylevel image into the wavelet representation of a lower-resolution multispectral image). In a concept, called ARSIS, the statistics of the detail coefficients are modelled before substitution.<sup>34</sup> In both cases, fusion as well as merging, using the simple combination rules, important information can be lost. In the case of fusion, other bands than the one containing the maximum can contribute to an improved visualization. In the case of merging, the low-resolution band can contain important directional information, that is not present in the substituted image.

Instead, we propose to use other rules, based on the proposed multispectral image wavelet representations. This concept allows for a detailed simultaneous description of directional information of all bands involved. The new representations are applied to combine a (high resolution) greylevel image with one band of a (low resolution) multispectral image. The obtained representations are reconstructed to replace the original band. This procedure is repeated for all the bands. It is demonstrated that this technique improves spatial resolution while maintaining the spectral resolution from the multispectral image.

The outline of the paper is as follows. In Sec. 2, the concept of Multiscale Fundamental Forms is introduced, leading to a redundant multispectral wavelet representation. In Sec. 3, the concept is extended towards orthogonal wavelet representations. In Sec. 4, we describe the application of multispectral image fusion and merging.

## 2. The redundant multispectral image wavelet representation

### 2.1. *Multispectral edge representation using the first fundamental form*

For the derivation of the first fundamental form, we will follow<sup>12</sup>. Let  $\mathbf{I}(x, y)$  be a multispectral image with domain  $\mathbb{R}^2$  and real-valued components  $I_n(x, y)$ ,  $n = 1, \dots, N$ . The value of  $\mathbf{I}$  at a given point is a  $N$ -dimensional vector. To describe the gradient information of  $\mathbf{I}$ , let us look at the differential of  $\mathbf{I}$ , which is assumed to exist. In the Euclidean space  $\mathbb{R}^N$ :

$$d\mathbf{I} = \frac{\partial \mathbf{I}}{\partial x} dx + \frac{\partial \mathbf{I}}{\partial y} dy \quad (1)$$

and its squared norm is given by (sums are over all bands of the image):



$$\begin{aligned}
\|d\mathbf{I}\|^2 &= \begin{pmatrix} dx \\ dy \end{pmatrix}^T \begin{pmatrix} \|\frac{\partial \mathbf{I}}{\partial x}\|^2 & \frac{\partial \mathbf{I}}{\partial x} \cdot \frac{\partial \mathbf{I}}{\partial y} \\ \frac{\partial \mathbf{I}}{\partial x} \cdot \frac{\partial \mathbf{I}}{\partial y} & \|\frac{\partial \mathbf{I}}{\partial y}\|^2 \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \\
&= \begin{pmatrix} dx \\ dy \end{pmatrix}^T \begin{pmatrix} \sum_n \left(\frac{\partial I_n}{\partial x}\right)^2 & \sum_n \frac{\partial I_n}{\partial x} \frac{\partial I_n}{\partial y} \\ \sum_n \frac{\partial I_n}{\partial x} \frac{\partial I_n}{\partial y} & \sum_n \left(\frac{\partial I_n}{\partial y}\right)^2 \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \\
&= \begin{pmatrix} dx \\ dy \end{pmatrix}^T \begin{pmatrix} G^{xx} & G^{xy} \\ G^{xy} & G^{yy} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \tag{2}
\end{aligned}$$

This expression is called the first fundamental form. The matrix

$$G = \begin{pmatrix} G^{xx} & G^{xy} \\ G^{xy} & G^{yy} \end{pmatrix} \tag{3}$$

is a  $2 \times 2$  symmetric, nonnegative definite matrix, which ensures that its eigenvalues are real and nonnegative. It reflects the change in a multispectral image. The direction of maximal and minimal change are given by the eigenvectors of  $G$ . The corresponding eigenvalues denote the rates of change. For a greylevel image ( $N = 1$ ), it is easily calculated that the largest eigenvalue is given by  $\lambda^+ = \|\nabla I\|^2$ , i.e. the squared gradient magnitude. The corresponding eigenvector lies in the direction of the gradient. The other eigenvalue  $\lambda^-$  equals zero. For a multispectral image, the eigenvectors and eigenvalues describe an ellipse in the image plane. When  $\lambda^+ \gg \lambda^-$ , the gradients of all bands are more or less in the same direction. When  $\lambda^- \simeq \lambda^+$ , there is no preferential direction, i.e. the eigenvectors can be chosen arbitrarily. The conjecture is that  $\lambda^-$  can be ignored and that the multispectral edge information is reflected by the first eigenvalue and corresponding eigenvector of the first fundamental form.

A particular problem that occurs is that the diagonalization does not uniquely specify the sign of the eigenvectors. This has been extensively studied<sup>13</sup>. It was proven that the eigenvectors can be uniquely oriented in simply connected regions where  $\lambda^- \neq \lambda^+$ . Based on this, an algorithm was proposed to orient the eigenvectors, keeping the angle-function continuous in local regions.

## 2.2. The dyadic wavelet transform

The first fundamental form of the previous section reflects only edge information at a single scale. Many greylevel image processing applications make use of higher scale edge information using multiresolution techniques.

The aim of this paper is to extend the multispectral edge representation towards multiscale edges.

The wavelet transform employed in this work is based on non-orthogonal (redundant) discrete wavelet frames introduced by Mallat.<sup>15,37</sup> Define a 2-D smoothing function  $\theta(x, y)$ , i.e.:

$$\begin{aligned} \lim_{x, y \rightarrow \infty} \theta(x, y) &= 0 \\ \int \theta(x, y) dx dy &= 1 \end{aligned} \quad (4)$$

Supposing  $\theta$  is differentiable, define

$$\psi^1(x, y) = \frac{\partial \theta(x, y)}{\partial x} \text{ and } \psi^2(x, y) = \frac{\partial \theta(x, y)}{\partial y} \quad (5)$$

The fact that  $\theta$  satisfies (4) guarantees that  $\psi^1(x, y)$  and  $\psi^2(x, y)$  are wavelets, i.e. have zero mean. The wavelets that were chosen in<sup>15</sup>, and that we also will apply, are quadratic spline wavelets of compact support, and  $\theta$  is a cubic spline smoothing function. The wavelet transform of an image  $I(x, y)$  is then defined by:

$$D_s^1(x, y) = I * \psi_s^1(x, y) \text{ and } D_s^2(x, y) = I * \psi_s^2(x, y) \quad (6)$$

where  $*$  denotes the convolution operator and

$$\psi_s^1(x, y) = \frac{1}{s^2} \psi^1\left(\frac{x}{s}, \frac{y}{s}\right) \text{ and } \psi_s^2(x, y) = \frac{1}{s^2} \psi^2\left(\frac{x}{s}, \frac{y}{s}\right) \quad (7)$$

denote the dilations of the functions  $\psi^i$ .  $s$  is the *scale* parameter which commonly is set equal to  $2^j$  with  $j = 1, \dots, d$ . This yields the so called *dyadic* wavelet transform of *depth*  $d$ .  $D_{2^j}^1$  and  $D_{2^j}^2$  are referred to as the detail images, since they contain horizontal and vertical details of  $I$  at scale  $j$ .

In practice, for digital images detail values are only required at integer positions  $(m, n) \in \mathbb{Z}^2$ . For finite image sizes, border problems are solved by imposing periodic boundary conditions. A fast algorithm has been designed for this transform by iterative filtering with a set of 1-dimensional low and high pass filters  $H, G, K$  and  $L$  associated with the wavelets  $\psi^1$  and  $\psi^2$  (for this, it is necessary that the wavelets can be written as separable products of 1-D functions).<sup>37</sup> These filters have finite impulse responses, because  $\psi^1$  and  $\psi^2$  have compact support, which makes the transform fast and easy to implement. We denote  $A_p$  the discrete filter obtained by putting  $2^p - 1$  zeros between each of the coefficients of the filter  $A$ . We denote by  $(A, B) * C$  the separable convolution of the rows and columns, respectively, of the discrete

image  $C$  with the discrete 1-D filters  $A$  and  $B$ .  $L_1 = I$  and  $D$  is the unit-impulse filter whose impulse response equals 1 at 0 and 0 otherwise. The procedure at scale  $j$  is given by:

$$\begin{aligned} L_{2^{j+1}}(m, n) &= (H_j, H_j) * L_{2^j}(m, n) \\ D_{2^{j+1}}^1(m, n) &= (D, G_j) * L_{2^j}(m, n) \\ D_{2^{j+1}}^2(m, n) &= (G_j, D) * L_{2^j}(m, n) \end{aligned} \quad (8)$$

For a proof, we refer to <sup>15</sup>. Thus the wavelet representation of depth  $d$  of the image  $I$  consists of the *low resolution image*  $L_{2^d}$  and *detail images*  $\{D_{2^j}^i\}_{j=1, \dots, d}^{i=1, 2}$ . Similarly, a reconstruction algorithm was designed, to reconstruct the original image from its wavelet representation. At scale  $j$ :

$$\begin{aligned} L_{2^{j-1}}(m, n) &= (K_{j-1}, L_{j-1}) * D_{2^j}^1(m, n) \\ &\quad + (L_{j-1}, K_{j-1}) * D_{2^j}^2(m, n) \\ &\quad + (\tilde{H}_{j-1}, \tilde{H}_{j-1}) * L_{2^j}(m, n) \end{aligned} \quad (9)$$

with  $\tilde{H}_p$  the filter whose fourier transform is the complex conjugate of the fourier transform of  $H_p$ .

Substitution of (5) and (7) in (6) yields the following interesting property:

$$\begin{pmatrix} D_{2^j}^1(x, y) \\ D_{2^j}^2(x, y) \end{pmatrix} = 2^j \begin{pmatrix} \frac{\partial}{\partial x}(I * \theta_{2^j})(x, y) \\ \frac{\partial}{\partial y}(I * \theta_{2^j})(x, y) \end{pmatrix} = 2^j \nabla(I * \theta_{2^j})(x, y) \quad (10)$$

This stipulates that the wavelet transform of a greylevel image consists of the components of the gradient of the image, smoothed by the dilated smoothing function  $\theta_{2^j}$ .

### 2.3. The multiscale fundamental form

Based on (10), for multispectral images a fundamental form can be constructed at each scale. Similar to (2), and applying (10), the squared norm of the differential of  $(\mathbf{I} * \theta_{2^j})(x, y)$  is given by:

$$\begin{aligned} & \|d(\mathbf{I} * \theta_{2^j})\|^2 \\ &= 2^{-2j} \begin{pmatrix} dx \\ dy \end{pmatrix}^T \begin{pmatrix} \sum_n (D_{n,2^j}^1)^2 & \sum_n D_{n,2^j}^1 D_{n,2^j}^2 \\ \sum_n D_{n,2^j}^1 D_{n,2^j}^2 & \sum_n (D_{n,2^j}^2)^2 \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \\ &= 2^{-2j} \begin{pmatrix} dx \\ dy \end{pmatrix}^T \begin{pmatrix} G_{2^j}^{xx} & G_{2^j}^{xy} \\ G_{2^j}^{xy} & G_{2^j}^{yy} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \end{aligned} \quad (11)$$

where  $D_{n,2^j}^1$  and  $D_{n,2^j}^2$  are the  $j$ -th scale detail coefficients of the  $n$ -th band image. The matrices

$$G_{2^j} = \begin{pmatrix} G_{2^j}^{xx} & G_{2^j}^{xy} \\ G_{2^j}^{xy} & G_{2^j}^{yy} \end{pmatrix} \quad (12)$$

are  $2 \times 2$  symmetric, nonnegative definite matrices, with real and non-negative eigenvalues. Expression (19) will be referred to as the  $j$ -th scale fundamental form. It reflects the change in the  $j$ -th scale smoothed image and therefore the edge information at the  $j$ -th scale. The direction of maximal and minimal change are given by the eigenvectors  $\mathbf{v}_{2^j}^+$  and  $\mathbf{v}_{2^j}^-$  of  $G$ . The corresponding eigenvalues  $\lambda_{2^j}^+$  and  $\lambda_{2^j}^-$  denote the rates of change. The eigenvectors and eigenvalues describe an ellipse in the image plane, where the longest axis denotes the direction of the gradient at scale  $j$  and the shortest axis the variance of gradient at scale  $j$  around that direction. For a greylevel image, one obtains

$$\begin{aligned} \lambda_{2^j}^+(x, y) &= 2^{-2j} [(D_{2^j}^1)^2(x, y) + (D_{2^j}^2)^2(x, y)] = \|\nabla(I * \theta_{2^j})\|^2 \\ \mathbf{v}_{2^j}^+(x, y) &= \frac{\nabla(I * \theta_{2^j})}{\|\nabla(I * \theta_{2^j})\|} \end{aligned} \quad (13)$$

i.e. the first eigenvector denotes the direction of the gradient of the  $j$ -th scale smoothed image, while its corresponding eigenvalue denotes its length. For  $N = 1$ , the matrices  $G_{2^j}$  have rank 1, i.e.  $\lambda_{2^j}^-(x, y) = 0$ . Remark that:

$$\begin{aligned} D_{2^j}^1(x, y) &= \sqrt{\lambda_{2^j}^+} v_{2^j,x}^+(x, y) \\ D_{2^j}^2(x, y) &= \sqrt{\lambda_{2^j}^+} v_{2^j,y}^+(x, y) \end{aligned} \quad (14)$$

i.e. the original representation is obtained in terms of the first eigenvalue and corresponding eigenvector. In multispectral images the edge information is contained in both eigenvalues. The eigenvectors and eigenvalues of the multiscale fundamental forms describe the edge information of a multispectral image in a multiresolution way. As in the single-scale case,  $\lambda_{2^j}^-$  is ignored. The multispectral image is then represented at each scale by 2 detail images:

$$\begin{aligned} D_{2^j}^{1,+}(x, y) &= \sqrt{\lambda_{2^j}^+} v_{2^j,x}^+(x, y) \\ D_{2^j}^{2,+}(x, y) &= \sqrt{\lambda_{2^j}^+} v_{2^j,y}^+(x, y) \end{aligned} \quad (15)$$

In Fig. 1, a schematic overview of the Multispectral Image Wavelet Representation (MIWR) is given. It is important to remark that this repre-

sensation is by no means complete, i.e. the original bands cannot be reconstructed from this representation. It is not the goal of this work to derive a complete representation but merely to describe the edge information of multispectral images.

The same problem as in the single-scale case occurs with calculating the MIWR: the matrix diagonalization does not uniquely specify the signs of the eigenvectors. This phenomenon translates in the multispectral image problem as arbitrariness of the gradients orientation. From (15), this orientation reflects on the sign of the detail coefficients that can flip incoherently from one pixel to another. Therefore the orientation must be determined before a reconstruction can be calculated. Instead of following the proposal of <sup>13</sup>, we propose the following (more simple) solution to this problem. The orientation of the gradient is approximated by the orientation of the gradient of the average of all bands. The average of the bands is calculated and wavelet transformed. The scalar product of the obtained detail coefficients  $\overline{D}_{2^j}^1$  and  $\overline{D}_{2^j}^2$  with the first eigenvectors then determines the signs: if  $\overline{D}_{2^j}^1 v_{2^j,x}^+ + \overline{D}_{2^j}^2 v_{2^j,y}^+ \geq 0$  then the sign of the eigenvectors is not changed, if the scalar product is negative, then the sign of  $\mathbf{v}_{2^j}^+$  is flipped. The sign of  $\mathbf{v}_{2^j}^-$  is chosen so that the angle of its direction is  $\frac{\pi}{2}$  more than the angle of  $\mathbf{v}_{2^j}^+$ .

### 3. The orthogonal multispectral image wavelet representation

#### 3.1. *Multispectral maximal length of linear vector operators in the image plane*

Define a vector operator  $\mathbf{L}$  in the  $(x, y)$ -plane, with components  $L_x$  and  $L_y$ , both linear operators. The components are taken to be identical operators, applied in 1D,  $L_x$  applied in the  $x$ -direction and  $L_y$  applied in the  $y$ -direction. Examples of such operators are the gradient operator  $\nabla = (\partial/\partial x, \partial/\partial y)$ , or any vector operator, consisting of two identical translation invariant convolution operators. When such an operator works on a scalar image  $I(x, y)$ , the result is a vector, with components  $L_x(I)$  and  $L_y(I)$ . This vector has length  $\sqrt{(L_x(I))^2 + (L_y(I))^2}$  and direction, given by  $(\cos \alpha, \sin \alpha)$ , with  $\alpha = \arctan(L_y(I)/L_x(I))$ .

How to apply such operator on a multispectral image? One obvious procedure would be to apply it on each band separately, and average the resulting vectors, i.e.  $\mathbf{L}(\mathbf{I}) = \frac{1}{N} \sum_n \mathbf{L}(I_n)$ . In other cases, one is interested

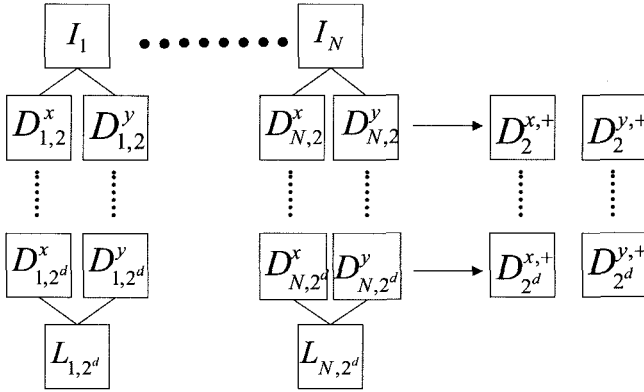


Fig. 1. Schematic overview of the Multispectral Image Wavelet Representation (MIWR)

in the maximum effect of the operator, e.g. in case of the gradient operator, this would lead to the most dominant edges. Therefore, the operator can be applied on each band separately, and the resulting vector with largest length:  $\mathbf{L}(\mathbf{I}) = \mathbf{L}(L_j); j = \arg \max_n \|\mathbf{L}(I_n)\|^2$  can be chosen. However, both procedures do not take the bands into account simultaneously. In case of the average, opposite resulting vectors will annihilate the effect of the operator. In case of the maximum, important 'next-to' largest effects can get lost.

In this section, a multispectral 'maximal length' and 'direction of maximal length' of vector operators is defined, taking into account all bands simultaneously. The procedure is derived from a concept similar to that of the First Fundamental Form in Sec. 2.

Define the infinitesimal displacement vector:

$d\mathbf{r} = (dx, dy)$ . Then, the scalar product in  $\mathbb{R}^2$ :  $\mathbf{L}(\mathbf{I}) \cdot d\mathbf{r} = L_x(\mathbf{I})dx + L_y(\mathbf{I})dy$ . Its squared norm in  $\mathbb{R}^N$ , using the Euclidean metric is given by:

$$\|\mathbf{L}(\mathbf{I}) \cdot d\mathbf{r}\|^2 = \begin{pmatrix} dx \\ dy \end{pmatrix}^T \begin{pmatrix} M^{xx} & M^{xy} \\ M^{xy} & M^{yy} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} \quad (16)$$

This expression is a quadratic form. The matrix

$$\begin{aligned}
 M &= \begin{pmatrix} \|L_x(\mathbf{I})\|^2 & L_x(\mathbf{I}) \cdot L_y(\mathbf{I}) \\ L_x(\mathbf{I}) \cdot L_y(\mathbf{I}) & \|L_y(\mathbf{I})\|^2 \end{pmatrix} \\
 &= \begin{pmatrix} \sum (L_x(I_n))^2 & \sum (L_x(I_n))(L_y(I_n)) \\ \sum (L_x(I_n))(L_y(I_n)) & \sum (L_y(I_n))^2 \end{pmatrix}
 \end{aligned} \tag{17}$$

(sums are over  $n = 1, \dots, N$ ) is a  $2 \times 2$  symmetric, nonnegative definite matrix, which ensures that its eigenvalues are real and nonnegative. For varying  $d\mathbf{r}$ , this form has a maximum and a minimum given by the largest eigenvalue  $\lambda^+$  and smallest eigenvalue  $\lambda^-$  of the matrix  $M$ . The maximal and minimal directions are given by the corresponding eigenvectors  $\mathbf{v}^+$  and  $\mathbf{v}^-$ .

For a greylevel image ( $N = 1$ ), it is easily calculated that the largest eigenvalue is given by  $\lambda^+ = \|\mathbf{L}(I)\|^2$  and  $\mathbf{v}^+ = (\cos \alpha, \sin \alpha)$ , with  $\alpha = \arctan L_y(I)/L_x(I)$ . The other eigenvalue  $\lambda^-$  equals zero. Therefore:

$$\mathbf{L}(I(x, y)) = \begin{pmatrix} \sqrt{\lambda^+} v_x^+(x, y) \\ \sqrt{\lambda^+} v_y^+(x, y) \end{pmatrix} \tag{18}$$

For a multispectral image, the second eigenvalue generally differs from zero. The first eigenvalue denotes the maximal squared length of the vector  $\mathbf{L}(\mathbf{I})$ , while the corresponding eigenvector lies in the direction of the maximal length. The second eigenvector lies in the opposite direction of maximal length. The corresponding eigenvalue denotes the minimal length. We will ignore the second eigenvalue and define:

$$\mathbf{L}(\mathbf{I}(x, y)) \equiv \begin{pmatrix} \sqrt{\lambda^+} v_x^+(x, y) \\ \sqrt{\lambda^+} v_y^+(x, y) \end{pmatrix} \tag{19}$$

The same problem as in Sec. 2 with the signs of the eigenvectors occurs, and the same solution is proposed. The scalar product of the two vectors  $\mathbf{L}(\mathbf{I})$  and  $1/N \sum_n \mathbf{L}(I_n)$  is evaluated. If the sign of the scalar product is positive, the direction of the eigenvectors is maintained, else it is flipped.

As already mentioned, the most obvious vector operator on which the proposed technique can be applied is the gradient operator, where the quadratic form is referred to as the first fundamental form. In Sec. 2, we have extended this concept to multiscale gradients, defining multiscale fundamental forms. One major drawback was that the applied wavelet transform was redundant, leading to large computer power and memory requirements.

With the proposed extension of the fundamental form towards general linear operators, an orthogonal multispectral wavelet representation can be constructed, making use of the DWT. In the next section, we will show how to develop such an orthogonal multispectral wavelet representation.

### 3.2. An orthogonal multispectral wavelet representation

In the dyadic separable orthogonal wavelet decomposition <sup>38</sup>, an image  $f(x, y)$  is decomposed at different scales  $j$  into a lower resolution image  $A_{2^j} f$  and three detail images  $D_{2^j}^1 f, D_{2^j}^2 f$  and  $D_{2^j}^3 f$ . This is done with a uniform sampling of two-dimensional convolution products. At each scale  $-1 \geq j \geq -J; (n, m) \in Z^2$ :

$$\begin{aligned}
 A_{2^j} f &= f(x, y) \otimes \phi_{2^j}(-x)\phi_{2^j}(-y)(2^{-j}n, 2^{-j}m) \\
 D_{2^j}^1 f &= f(x, y) \otimes \phi_{2^j}(-x)\psi_{2^j}(-y)(2^{-j}n, 2^{-j}m) \\
 D_{2^j}^2 f &= f(x, y) \otimes \psi_{2^j}(-x)\phi_{2^j}(-y)(2^{-j}n, 2^{-j}m) \\
 D_{2^j}^3 f &= f(x, y) \otimes \psi_{2^j}(-x)\psi_{2^j}(-y)(2^{-j}n, 2^{-j}m)
 \end{aligned}
 \tag{20}$$

Here,  $\phi$  and  $\psi$  are perfect low and bandpass filters, that are separably applied along  $x$  and  $y$ -directions. This wavelet representation can be computed in practice with a pyramidal algorithm, making use of 1D quadrature mirror filters  $H$  (lowpass) and  $G$  (bandpass), that are applied separably in the  $x$  and  $y$ -directions. In Fig. 2, the algorithm is illustrated by a block diagram for one scale. The process consists out of separable 1-D convolutions along row (column) and subsampling, by skipping one row (column) out of two.

The wavelet representation is complete, and reconstruction of the original image is accomplished by putting zero's between each sample of the detail images and convolution with filters  $\hat{H}$  and  $\hat{G}$ , which are symmetric filters, with impulse responses, given by  $\hat{h}(n) = h(-n)$  and  $\hat{g}(n) = g(-n)$ . Again, for images this is separably applied on the rows and columns. In Fig. 3, the reconstruction scheme is illustrated by a block diagram for one scale.

A representation for multispectral images will be developed by combining detail information of the different bands into one set of detail images. For this, let us define the following linear vector operator:

$$\mathbf{L} = \begin{pmatrix} G_x \\ G_y \end{pmatrix} \otimes \tag{21}$$



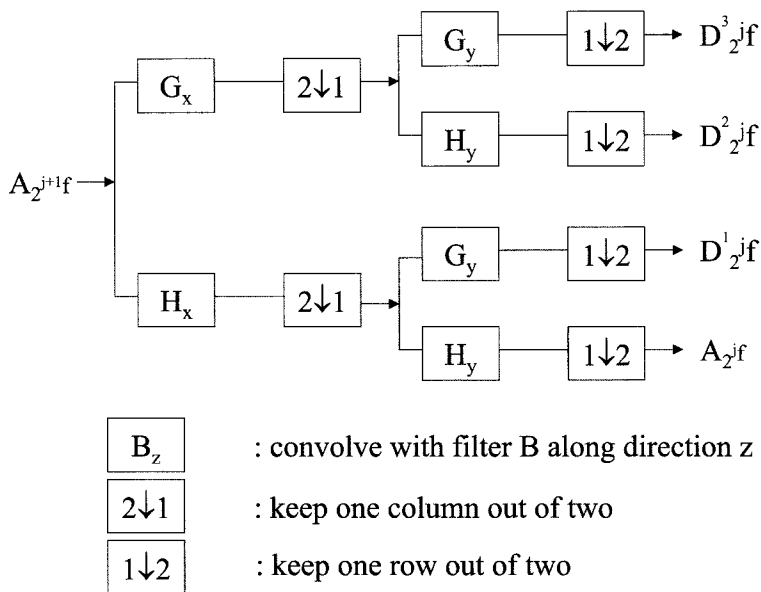


Fig. 2. the pyramidal orthogonal wavelet decomposition

i.e. an operator, with components the bandpass filters  $G_x$  and  $G_y$ . This operator is applied on a multispectral image by using (19).

Actually,  $L$  is the first operator that is applied within the pyramidal decomposition. In order to see this, the following modification is required (see Fig. 4). In the modified algorithm, the order of applying  $H_x$  and  $G_y$  is reversed when calculating  $D_{2^j}^1 f$ . In this way, the bandpass filter is applied first in  $x$  and  $y$  directions. Remark that this modification is allowed, since it only changes the order of convolution with  $\psi$  and  $\phi$  for calculating  $D_{2^j}^1 f$  in (20). Remark also that the calculation of the  $D^3$  images can be done in two ways, leading to exactly the same results.

The multispectral image wavelet representation at each scale is defined in three steps. In the first step, the lowpass images  $A_{2^{j+1}} I_n$  of the bands  $I_n, n = 1, \dots, N$  are bandpass filtered with  $G_x$  and  $G_y$  (see Fig. 5), leading to detail images  $D_{2^j}^x I_n$  and  $D_{2^j}^y I_n$ . The operator  $L$  is defined as in (21), which means that in fact:

$$\begin{pmatrix} D_{2^j}^x I_n \\ D_{2^j}^y I_n \end{pmatrix} = \begin{pmatrix} L_x A_{2^{j+1}} I_n \\ L_y A_{2^{j+1}} I_n \end{pmatrix} \quad (22)$$

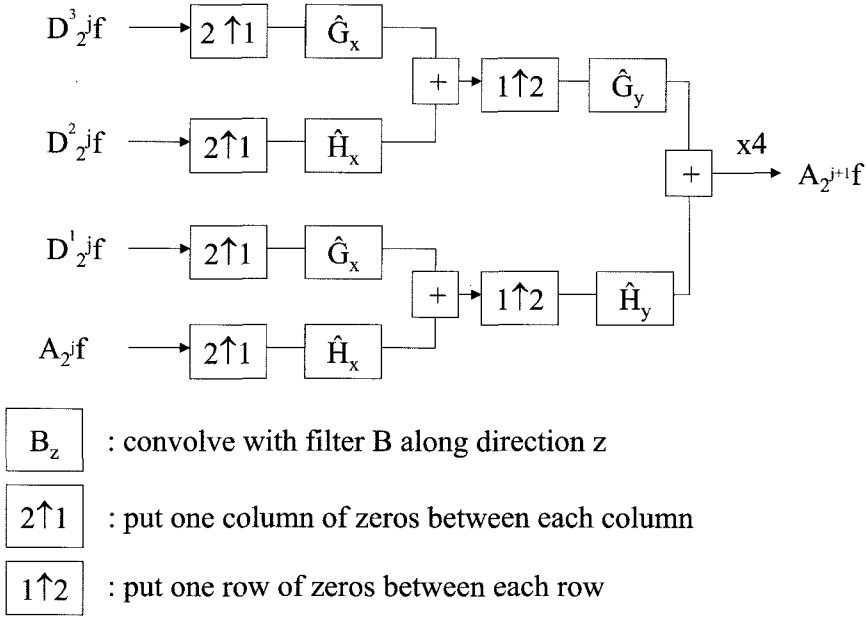


Fig. 3. the pyramidal orthogonal wavelet reconstruction

In the second step, we calculate  $\mathbf{L}(\mathbf{A}_{2^{j+1}}\mathbf{I}) = \mathbf{L}(A_{2^{j+1}}I_1, A_{2^{j+1}}I_2, \dots, A_{2^{j+1}}I_N)^T$  To do so, the matrix:

$$\begin{aligned}
 & \begin{pmatrix} \sum (L_x A_{2^{j+1}} I_n)^2 & \sum L_x A_{2^{j+1}} I_n L_y A_{2^{j+1}} I_n \\ \sum L_x A_{2^{j+1}} I_n L_y A_{2^{j+1}} I_n & \sum (L_y A_{2^{j+1}} I_n)^2 \end{pmatrix} \\
 &= \begin{pmatrix} \sum (D_{2^j}^x I_n)^2 & \sum (D_{2^j}^x I_n)(D_{2^j}^y I_n) \\ \sum (D_{2^j}^x I_n)(D_{2^j}^y I_n) & \sum (D_{2^j}^y I_n)^2 \end{pmatrix}
 \end{aligned} \tag{23}$$

is diagonalized, leading to eigenvalues  $\lambda_{2^j}^+$  and  $\lambda_{2^j}^-$ , with corresponding eigenvectors  $\mathbf{v}_{2^j}^+$  and  $\mathbf{v}_{2^j}^-$ . Then, according to (19),

$$\mathbf{L}\mathbf{A}_{2^{j+1}}\mathbf{I} = \begin{pmatrix} \sqrt{\lambda_{2^j}^+} \mathbf{v}_{2^j}^{+,x} \\ \sqrt{\lambda_{2^j}^+} \mathbf{v}_{2^j}^{+,y} \end{pmatrix} \equiv \begin{pmatrix} D_{2^j}^{+,x} \mathbf{I} \\ D_{2^j}^{+,y} \mathbf{I} \end{pmatrix} \tag{24}$$

Thirdly, the images  $D_{2^j}^x I_n$  and  $D_{2^j}^y I_n$  are replaced by  $D_{2^j}^{+,x} \mathbf{I}$  and  $D_{2^j}^{+,y} \mathbf{I}$ . These detail images are then treated as being detail images from a represen-

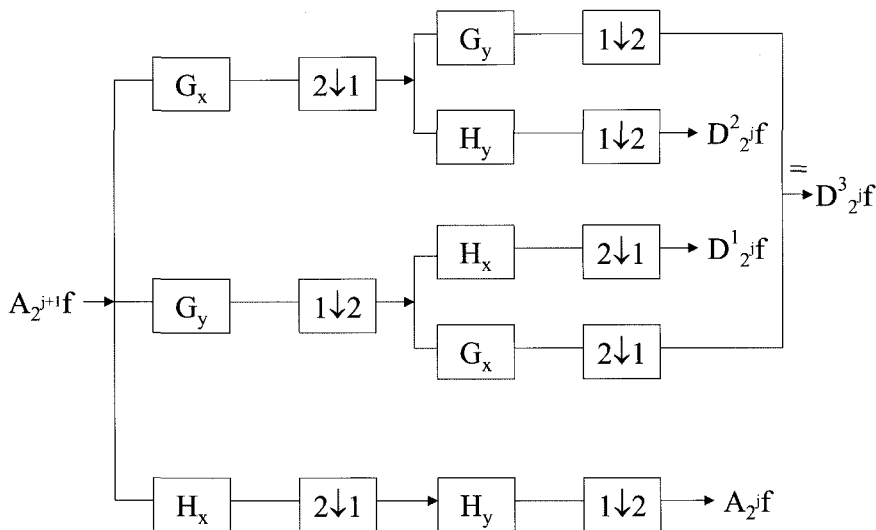


Fig. 4. a modified pyramidal orthogonal wavelet decomposition

tation of one greylevel image. Then, the remaining of the pyramidal algorithm of Fig. 2 is calculated (see Fig. 6). This leads to detail images  $D_{2^j}^{+,1}\mathbf{I}$ ,  $D_{2^j}^{+,2}\mathbf{I}$  and  $D_{2^j}^{+,3}\mathbf{I}$ . Remark that the two ways of calculating  $D_{2^j}^{+,3}\mathbf{I}$  generally will lead to (slightly) different results. In the following, we will take the average of both results. Also remark that the low resolution images  $A_{2^j} I_n$  are calculated for each band separately. After decomposition, the following images are generated: one set of detail images  $\{D_{2^j}^{+,1}\mathbf{I}, D_{2^j}^{+,2}\mathbf{I}, D_{2^j}^{+,3}\mathbf{I}\}_{-1 \geq j \geq -J}$  and for each band the lowest resolution images  $A_{2^j} I_n$ .

In this way, an orthogonal wavelet representation of multispectral images is obtained, similar as the redundant representation of Sec. 2, but with the advantage of reduced computational complexity and memory requirements. Reconstruction can be accomplished by using the pyramidal reconstruction scheme of Fig. 3, with the detail images, obtained by the presented decomposition scheme.

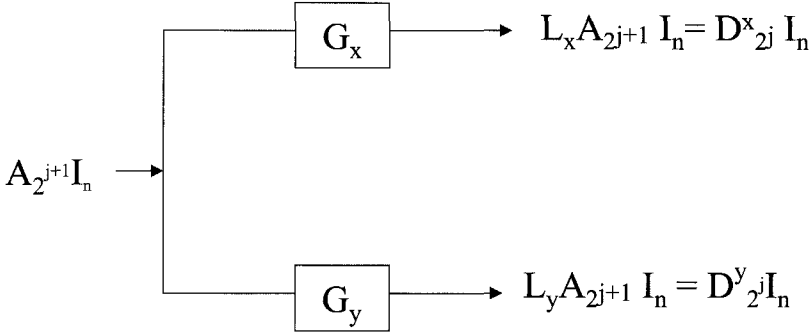


Fig. 5. the first step: bandpass filtering

## 4. Applications

### 4.1. Multispectral image fusion

In this work, the redundant representation as well as the orthogonal representation of multispectral images are applied for fusion. Using the proposed redundant representation, a fusion algorithm can be constructed in the following way. All bands are wavelet transformed using (8). For each scale, the multiscale fundamental forms are calculated using (19). After diagonalization, the wavelet representation (15) is obtained. A low resolution image is obtained by averaging the low resolution images of the original bands:  $\bar{L}_{2^d} = \frac{1}{N} \sum_{n=1}^N L_{n,2^d}$ . The obtained representation is then given by:  $\bar{L}_{2^d}$  and  $\{D_{2^j}^{i,+}\}_{j=1,\dots,d}$ . This representation is reconstructed by inverse wavelet transformation.

Using the proposed orthogonal representation, a fusion algorithm is constructed in the following way. First, all bands are transformed using the first step of the modified transform (see Fig. 5). The obtained detail images  $D_{2^j}^x I_n$  and  $D_{2^j}^y I_n$  are combined into  $D_{2^j}^{+,x} \mathbf{I}$  and  $D_{2^j}^{+,y} \mathbf{I}$  by diagonalizing (23). On these detail images, the remaining of the transform is applied (see Fig. 6). The low resolution images of all bands are averaged. The obtained representation is given by:  $\{D_{2^j}^{+,1} \mathbf{I}, D_{2^j}^{+,2} \mathbf{I}, D_{2^j}^{+,3} \mathbf{I}\}_{-1 \geq j \geq -J}$  and the lowest resolution image  $\langle A_{2^j} I_n \rangle$ . Reconstruction is performed, as in Fig. 5.

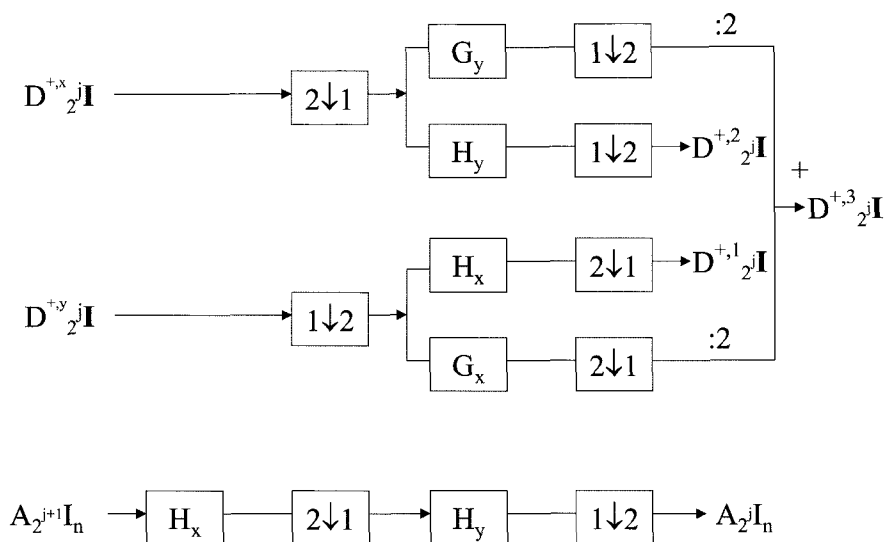


Fig. 6. the remaining of the pyramidal decomposition

To demonstrate the proposed fusion technique, the following experiment is conducted. As a test image remote sensing data is used: a Thematic Mapper image, containing 7 bands of 512x512 images from the U.S. Landsat series of satellites<sup>39</sup>. The first four images are fused into one greylevel image. In Fig. 7, the result using the redundant wavelet representation is shown. On the left, the proposed technique is applied. On the right, the wavelet fusion technique of<sup>22</sup> is applied where the same redundant wavelet transform is applied. In both applications, the images were wavelet decomposed into 4 levels of resolution. On the right image, for each pixel position and at each scale, the largest detail coefficient is taken to be the detail coefficient of the fused image:  $\tilde{D}_{2^j}^i(x, y) = \max_n D_{n, 2^j}^i(x, y)$ . One can observe an improved overall contrast using the proposed technique. This effect can be attributed to the superior description of the edge information in the MIWR representation. The same experiment was repeated using the orthogonal representation, with results that did not differ visually from Fig. 7.

At first sight, it might seem surprising that the fused images are reconstructed without visual artifacts. After all, the detail images  $\{D_{2^j}^{i,+}\}$  are

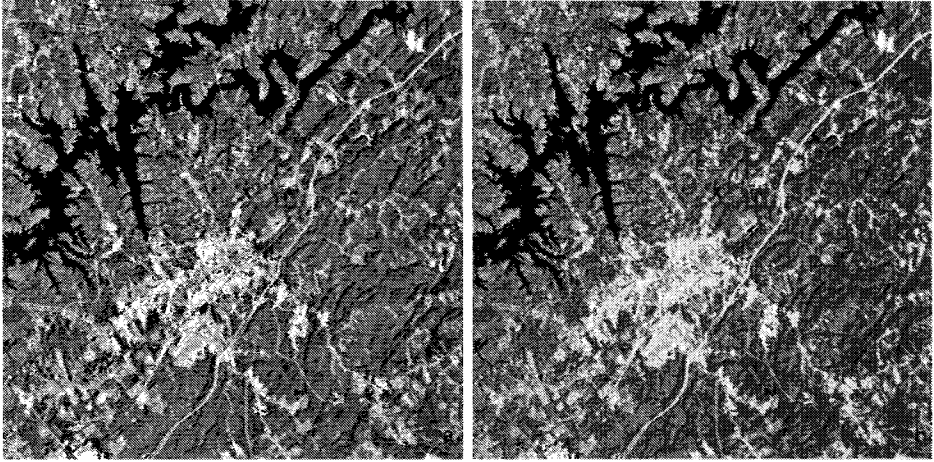


Fig. 7. Fusion of the first 4 bands of a Landsat image; a: using the proposed redundant representation; b: using wavelet maxima fusion

constructed pixelwise, irrespective of detail values at neighbouring positions and scales. This seems not to affect the reconstruction. We are not able to give rigorous mathematical evidence for this, but several arguments are listed below.

- For most positions,  $\lambda_{2^j}^-$  are small compared to  $\lambda_{2^j}^+$ . This is the case because most edge positions appear in more than one band and are equally oriented. If edge positions of different bands are equally strong, they are fused accordingly. If an edge position of one of the bands is stronger than the others, it will dominate the fusion process. In those cases where  $\lambda_{2^j}^- \simeq \lambda_{2^j}^+$ , the gradients are mostly weak. These positions correspond to noisy and locally textured areas.
- Real edges have continuous behaviour both in space and scale. This means that a strong gradient at one position at a particular scale will also be found at the same or a neighbouring position at the other scales, and this will be also the case for all positions along the edge. This is probably the main reason for the fused reconstruction to work properly. In <sup>15</sup>, a thorough study of the behaviour of edges accross scales is performed, revealing e.g. that an image can almost perfectly be reconstructed from the local maxima of the detail images only.
- Most of the other wavelet-based fusion techniques that have been

described in the literature, all are based on fusion rules that do not take into account neighbouring detail information in space or scale. No visual artifacts of the reconstructions were reported in these cases. The continuity arguments are also valid in those cases. The fact that the proposed technique combines gradient information of the different bands is an extra argument for a more continuous behaviour of the values of  $\{D_{2^j}^{i,+}\}$  across space and scale.

This experiment merely shows that the proposed fusion technique visually leads to similar results as the wavelet-based fusion techniques from the literature. The proposed technique appears to have an improved overall contrast compared to the wavelet maxima procedure, but it is hard to quantify these results. For this, human observer experiments should be performed. For fusion, particular task performance has been studied.<sup>40,41,42,43,44</sup> In the next section, we will demonstrate using specific task performance that the proposed technique is extremely useful for a specific fusion process, namely the merging of multispectral images.

#### 4.2. *Multispectral image merging*

In this section, we will discuss an application, closely related to fusion, which is the process of merging of different images. By merging we mean the exchange of useful information between two (or more) images, with the purpose of enhancing one or both images. One can e.g. merge a monochromatic image of high spatial resolution with a multispectral image of lower spatial resolution, with the purpose of enhancing the spatial resolution of the multispectral image. Applications are found in satellite imagery, where e.g. a monochromatic SPOT image (10 m resolution) can be merged with a 7-band LANDSAT image (30 m resolution).

Most merging techniques from the literature are developed within a multiresolution framework. When using the wavelet transform, merging is performed by wavelet transforming the monochromatic image and all bands of the multispectral image. A standard merging rule is to replace the detail images from the multispectral bands by the detail images from the monochromatic image.

Let us first introduce the merging procedure based on the presented representation. Suppose that we want to merge a greylevel image  $f(x, y)$  and a multispectral image  $\mathbf{g}(x, y) = (g_1(x, y), g_2(x, y), \dots, g_N(x, y))^T$ . The merging will be performed by combining the greylevel image and one of the bands of the multispectral image with the presented representation, to

obtain an enhanced band. This is then repeated for all of the bands of the multispectral image. In concreto, the following procedure is performed:

for  $i = 1$  to  $N$

- Regard the greylevel image  $f(x, y)$  and the image  $g_i(x, y)$  as the two bands of a 2-valued image  $\mathbf{I}$ , and calculate the presented representation.
- Reconstruct the merged image  $\hat{g}_i(x, y)$ , using its original lowest resolution image  $A_{2^j} g_i$  and the obtained detail images from the representation  $D_{2^j}^{+,1} \mathbf{I}$ ,  $D_{2^j}^{+,2} \mathbf{I}$  and  $D_{2^j}^{+,3} \mathbf{I}$ .

In this way, a merged multispectral image is obtained that contains combined detail information from the greylevel image and the original multispectral image. This information may be high spatial resolution information from the greylevel image or local resolution information from the specific bands of the multispectral image. Two goals are pursued: the improvement of the spatial resolution from the multispectral image, which is accomplished by including detail information from the greylevel image, and preservation of the spectral information from the multispectral image. The latter is accomplished by retaining the band at the coarsest scale and by including local detail information from the multispectral image, when not present in the greylevel image.

As an application, merging is applied to multispectral satellite imagery. The proposed merging technique will be applied. To compare, the substitution techniques from the literature are applied. Three different techniques are used. In the first, the detail images of each band of the multispectral image are replaced by the detail images of the panchromatic image.<sup>30</sup> We will refer to this technique as MERGE1. In the second approach, the same replacement takes place, but on top of this, the low-resolution wavelet images of each band of the multispectral image are replaced by the original bands.<sup>31</sup> We will refer to this technique as MERGE2. We also compare with a standard merging method, based on the Intensity-Hue-Saturation transformation.<sup>45</sup> Here, the multispectral image is IHS-transformed, and the panchromatic image is merged into the I-component. We will refer to this technique as IHS.

In the experiment, A SPOT Panchromatic image is merged with a three-band Landsat multispectral image. The SPOT image has a resolution of  $10m$ , while the Landsat images have a resolution of  $30m$ . In Fig. 8, the original SPOT and Landsat images are displayed. In Fig. 9, the merged results are shown, using IHS (a), MERGE1 (b), MERGE2 (c) and the



merged result using the proposed technique (d).



Fig. 8. a: Original SPOT and b: Landsat images

With respect to the spatial resolution, results are visually not really different. All four techniques clearly improve the spatial resolution of the multispectral image. With respect to the spectral resolution the results are visually different. IHS, MERGE1 and MERGE2 display a poor spectral resolution. The spectral information of the proposed technique very much resembles that of the original Landsat image. In most of the image only small color differences are visible, while at small specific area's, the color differences become large. To quantify the results, a correlation measure is applied.

The correlation between 2 images  $A$  and  $B$  is defined as:

$$Cor(A, B) = \frac{\langle (A - \langle A \rangle)(B - \langle B \rangle) \rangle}{\sqrt{\langle (A - \langle A \rangle)^2 \rangle \langle (B - \langle B \rangle)^2 \rangle}} \quad (25)$$

where  $\langle . \rangle$  denotes the average over all pixels. The correlation between original and merged multispectral images is calculated for each band separately. Although it is not really clear whether this metric has any relation with visual perception, it is used regularly for multispectral imagery. Recently, there have been some attempts to include perception-based metrics.<sup>23,46,47</sup>

In table 1, correlation results are shown. It is clear that all wavelet-based techniques better preserve spectral information than the IHS technique. The obtained values from the IHS technique and the wavelet techniques



Fig. 9. Merged images from Fig. 3, using a: IHS; b: MERGE1; c: MERGE2; d: proposed technique

from the literature agree with perviously reported results.<sup>30</sup> The proposed technique outperforms the other two wavelet mergers.

	IHS	MERGE1	MERGE2	Proposed
R	0.69	0.81	0.78	0.92
G	0.69	0.83	0.80	0.92
B	0.66	0.78	0.76	0.90

Spectral preservation is an important issue, not only for visual purposes,

but also for specific task performance. Many applications perform classification based on the spectral information. It is clear from the images and from the correlation measure that the proposed technique will outperform the others with respect to classification. To show this, we measured the average spectral response of a small homogeneous green area in the original Landsat image (pointed to by an arrow in Fig. 8b, and the same area in the merged images. The (Euclidean) distance in the RGB-space between the cluster centers of the original and the merged results were 269, 180, 169 and 74 for IHS, MERGE1, MERGE2 and the proposed technique. Similar experiment were performed at other regions, leading to similar results.

Finally, the following classification experiment is performed. The original Landsat image is segmented by clustering its RGB-space. For this, we applied the k-means clustering algorithm, with  $k = 4$ . The pixels, belonging to one of the clusters are shown in Fig. 10a, revealing the objects that have a spectral response, corresponding to that specific cluster. In Fig. 10b and c, we measure the same spectral response (i.e. display all the pixels that belong to the same cluster), on the merged images, using MERGE2 and the proposed technique respectively. One can notice that most of the objects have disappeared when using MERGE2, while most of the objects have been classified using the proposed technique, due to its ability to preserve spectral resolution. Moreover, the objects clearly have improved in spatial resolution.

The proposed merging technique will apply to images with different variations of spatial and spectral resolution. As an example, we apply it to a hyperspectral image (220-band AVIRIS)(see Fig. 11a). Since there was no high-resolution panchromatic image available, we simulate one. As the original panchromatic image, the average of all bands is calculated (Fig. 11b). To obtain the low-resolution original hyperspectral image, all bands are blurred (averaged over 3x3 pixels)(Fig. 11c). Then, the proposed merging procedure was applied (Fig. 11d). One can notice that the spatial resolution of the hyperspectral image is improved by the merging procedure.

## 5. Conclusions

We have proposed 2 new wavelet representations for multispectral images. A redundant representation using dyadic wavelet frames is based on the concept of multiscale fundamental forms, a multiresolution extension of the first fundamental form, that describes edge information of multispectral images. An orthogonal representation based on the Discrete Wavelet

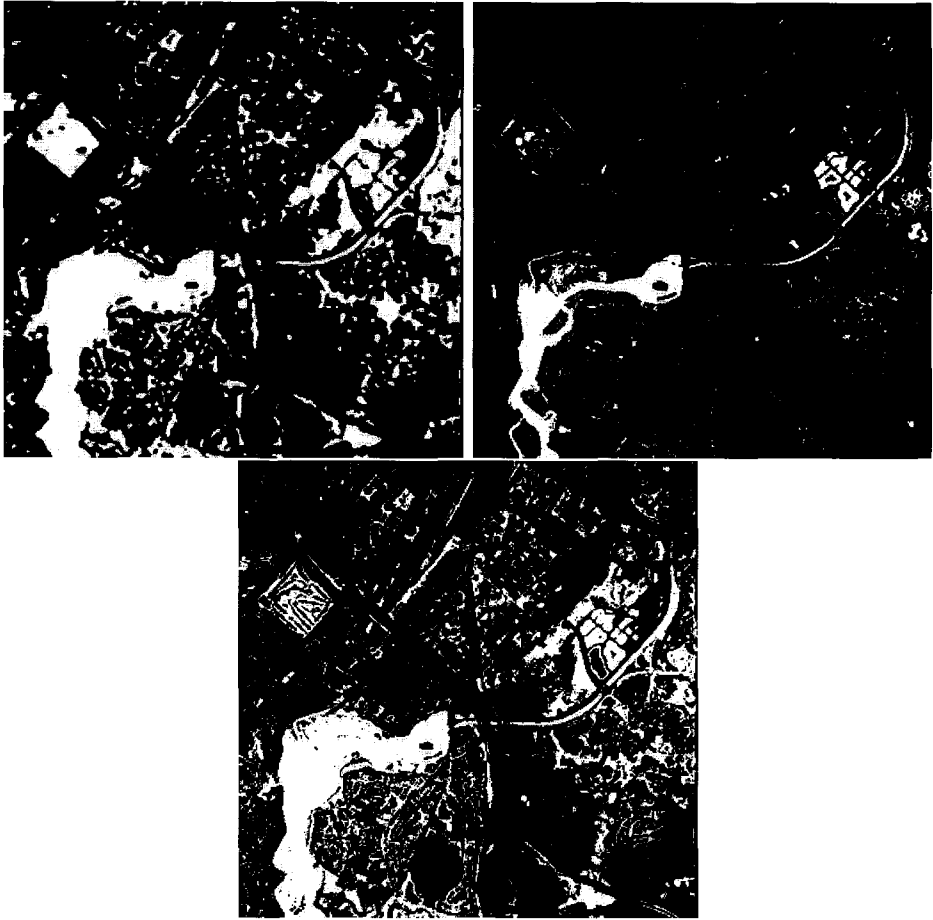


Fig. 10. Spectral response after k-means clustering a: on the original Landsat image; b: on the merged image, using MERGE2; c: on the merged image, using the proposed technique.

Transform is obtained by extending the concept of multiscale gradients to any linear vector operator in the image plane, having 2 identical components. The representations combine detail information from all the bands into one representation.

Based on the representation, multispectral image fusion and image merging techniques are proposed. Experiments are conducted for fusion and merging of multispectral satellite images. A Landsat TM image is fused and merged with a SPOT panchromatic image. The proposed techniques are demonstrated to outperform other wavelet-based merging techniques. Finally the merging technique is demonstrated on a hyperspectral image.

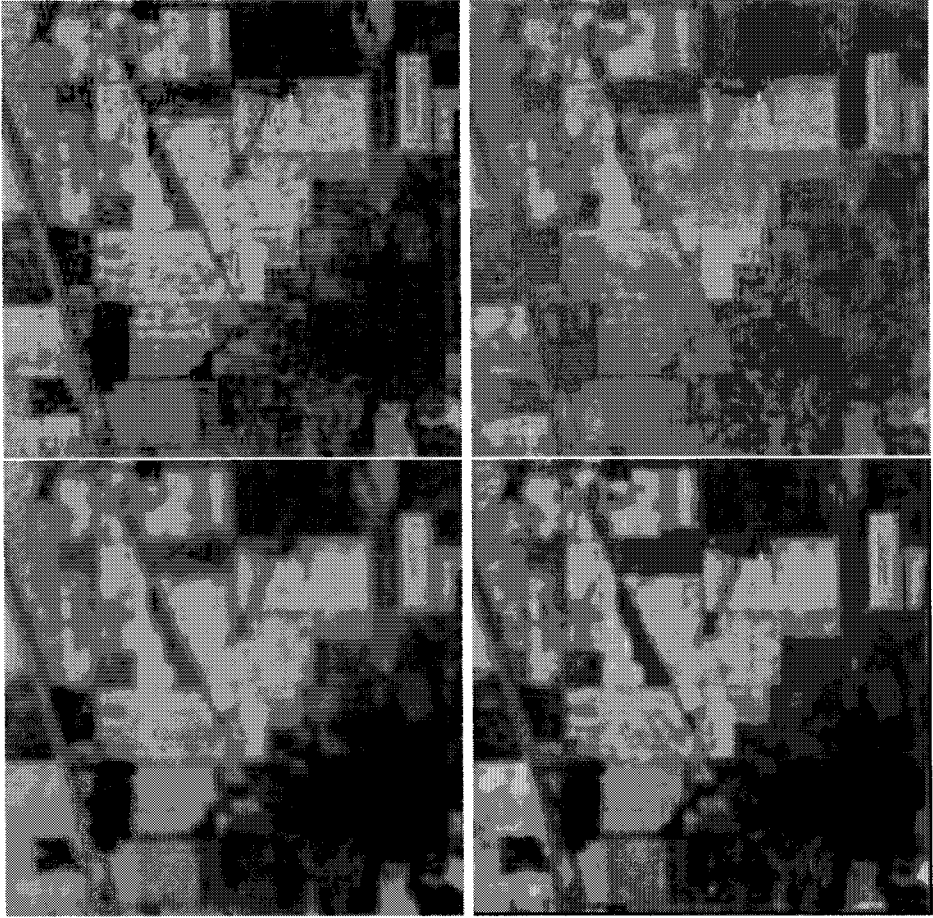


Fig. 11. Hyperspectral image merging; a: original hyperspectral image, band 170; b: panchromatic image; c: low-resolution hyperspectral image, band 170; d: merged hyperspectral image, band 170.

Other possible applications do not reconstruct the obtained multispectral representations, but only make use of the accumulated detail information. Applications include: anisotropic filtering of color and multispectral images<sup>18</sup>, wavelet-based denoising of multispectral images<sup>20</sup>, edge-based (e.g. watershed) multispectral image segmentation.<sup>19</sup>

## References

1. I.L. Thomas, V.M. Benning and N.P. Ching, Classification of remotely sensed images, Adam Hilger, Bristol (1987).
2. C. Lee and D.A. Landgrebe, Analyzing high-dimensional multispectral data,



- IEEE Trans. Geosci. Remote Sensing*, **31**:(4), pp. 388–400 (1993).
3. G. Harikumar and Y. Bresler, Multi-spectral image visualization by projection pursuit, *Visual. Biomed. Comput '92*, **SPIE-1808**, pp. 159–170 (1992).
  4. A. Manduca, Multispectral image visualization with nonlinear projections, *IEEE Transactions on Image Processing*, **5**:(10), pp. 1486–1490 (1996).
  5. D.A. Yocky, Image merging and data fusion by means of the discrete two-dimensional wavelet transform, *J. Opt. Soc. Am. A*, **12**:(9), pp. 1834–1841 (1995).
  6. S.T. Acton and J. Landis, Multispectral anisotropic diffusion, *Int. J. Remote Sensing*, **18**:(13), pp. 2877–2886 (1997).
  7. J.L. Starck and P. Querre, Multispectral data restoration by the wavelet karhunen-loeve transform, *Signal Processing*, **81**:(12), pp. 2449–2459 (2001).
  8. B. Aiazzi, L. Aparone and A. Barducci, Estimating noise and information of multispectral imagery, *Opt. Engineering*, **41**:(3), pp. 656–668 (2002).
  9. P.A. Mlsna and J.J. Rodriguez, A multivariate contrast enhancement technique for multispectral images, *IEEE Trans. Geosci Remote Sensing*, **33**:(1), pp. 212–216 (1995).
  10. J. Vrabel, Multispectral imagery band sharpening study, *Photogramm. Eng. Remote Sensing*, **62**:(9), pp. 1075–1083 (1996).
  11. M.J. Carlotte, Spatial enhancement of elevation data using a single multispectral image, *Opt. Engineering*, **39**:(2), pp. 430–437 (2000).
  12. S. Di Zenzo, A note on the gradient of a multi-image, *Comput. Vision Graphics Image Processing*, **33**, pp. 116–125 (1986).
  13. A. Cumani, Edge detection in multispectral images, *CVGIP: Graphical Models and Image Processing*, **53**, pp. 40–51 (1991).
  14. G. Sapiro and D.L. Ringach, Anisotropic diffusion of multivalued images with applications to color filtering, *IEEE Trans. Image Processing*, **5**:(11), pp. 1582–1586 (1996).
  15. S. Mallat and S. Zhong, Characterization of signals from multiscale edges, *IEEE Trans. Pattern Anal. Machine Intell.*, **14**, pp. 710–732 (1992).
  16. P. Scheunders, A multivalued image wavelet representation based on multiscale fundamental forms, *IEEE Transactions on Image Processing*, **11**:(5), pp. 568–575 (2002).
  17. P. Scheunders and S. De Backer, Fusion and merging of multispectral images with use of multiscale fundamental forms, *Journal of the Optical Society of America, A*, **18**:(10), pp. 2468–2477 (2001).
  18. P. Scheunders and J. Sijbers, Multiscale anisotropic filtering of color images, *Proceedings IEEE International Conference on Image Processing*, **III**, pp. 170–173 (2001).
  19. P. Scheunders and J. Sijbers, Multiscale watershed segmentation of multivalued images, *Proceedings of ICPR, International Conference on Pattern Recognition*, **III** (2002).
  20. P. Scheunders, Wavelet-based enhancement and denoising using multiscale structure tensor, *Proceedings of IEEE ICIP, International Conference on Image Processing*, to appear (2002).

21. C. Pohl and J. Van Genderen, Multisensor image fusion in remote sensing: concepts, methods and applications, *Int. J. Remote Sensing*, **19**, pp. 823–854 (1998).
22. H. Li, B.S. Manjunath and S.K. Mitra, Multisensor image fusion using the wavelet transform, *Graphical Models and Image Processing*, **57**:(3), pp. 235–245 (1995).
23. T. Wilson, S. Rogers and L. Meyers, Perceptual-based hyperspectral image fusion using multiresolution analysis, *Optical Engineering*, **34**:(11), pp. 3154–3164 (1995).
24. T. Wilson, S. Rogers and L. Meyers, Perceptual-based image fusion for hyperspectral data, *IEEE T. Geosci. Rem. Sensing*, **35**:(4), pp. 1007–1017 (1997).
25. T. Pu and G. Ni, Contrast-based image fusion using the discrete wavelet transform, *Optical Engineering*, **39**”(8), pp. 2075–2082 (2000).
26. Y. Carls-Powell, Fusion CCD and IR images creates color night vision, *Laser Focus World*, **32**:(5), pp. 32–36 (1996).
27. N. Nandhakumar and Y.K. Aggarwal, Integrated analysis of thermal and visual images for scene interpretation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **10**:(4), pp. 469–481 (1988).
28. B. GarguetDuport, The use of multiresolution analysis and wavelets transform for merging spot panchromatic and multispectral image data, *Photogramm. Eng. Rem. S.*, **62**:(9), pp. 1057–1066 (1996).
29. B. GarguetDuport, Wavemerg: A multiresolution software for merging spot panchromatic and spot multispectral data, *Environmental Modelling and Software*, **12**:(1), pp. 85–97 (1997).
30. D.A. Yocky, Multiresolution wavelet decomposition image merger of landsat thematic mapper and spot panchromatic data, *Photogramm. Engin. and Remote Sensing*, **62**, pp. 1067–1074 (1996).
31. J. Zhou, D. Civco and J. Silander, A wavelet transform method to merge landsat tm and spot panchromatic data, *Intern. J. Rem. Sensing*, **19**:(4), pp. 743–757 (1998).
32. J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala and R. Arbiol, Image fusion with additive multiresolution wavelet decomposition; applications to spot+landsat images, *J. Opt. Soc. Am. A*, **16**, pp. 467–474 (1999).
33. J. Nunez, X. Otazu, O. Fors, A. Prades, V. Pala and R. Arbiol, Multiresolution-based image fusion with additive wavelet decomposition, *IEEE T. Geosci. Rem. Sensing*, **37** (3), pp. 1204–1211 (1999).
34. T. Ranchin and L. Wald, Fusion of high spatial and spectral resolution images: the ARSIS concept and its implementation, *Photogramm. Engin. and Remote Sensing*, **66**, pp. 49–61 (2000).
35. F. Jahard, D.A. Fish, A.A. Rio and C.P. Thompson, Far/near infrared adapted pyramid-based fusion for automotive night vision, *Sixth International Conference on Image Processing and its Applications*, pp. 886–890 (1997).
36. B. Ajazzi, L. Alparone, S. Baronti and R. Carla Assessment of pyramid-based multisensor image data fusion, *Image and Signal Processing for Re-*

- mote Sensing*, S.B. Serpico, Ed., 1998, **IV**, pp. 237–248 (1998).
37. S. Mallat, *A wavelet tour of signal processing*, 2-nd edition, Academic Press (1999).
  38. S. Mallat, A theory for multiresolution signal decomposition : the wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**:(7), pp. 674–693 (1989).
  39. The images were provided by dr. C.C. Hung from the Center for Hydrology, Soil Climatology, and Remote Sensing from the Alabama A&M University.
  40. D. Ryan and R. Tinkler, Night pilotage assessment of image fusion, *SPIE Proceedings on Helmet and Head Mounted Displays and Symbology Design Requirements II*, Lewadowsky, R.J., Stephens, W., and Haworth, L.A., Eds., pp. 50–67 (1995).
  41. P.M. Steele and P. Perconti, Part task investigation of multispectral image fusion using gray scale and synthetic color night vision sensor imagery for helicopter pilotage, *Proceedings of the SPIE Conference on Targets and Backgrounds, Characterization and Representation III*, Watkins, W. and Clement, D., Eds., pp. 88–100 (1997).
  42. E.A. Essock, M.J. Sinai, J.S. McCarley, W.K. Krebs and J.K. Deford, Perceptual ability with real-world nighttime scenes: image-intensified, infrared and fused-color imagery, *Human Factors*, **41**:(3), pp. 438–452 (1999).
  43. M. Aguilar, D.A. Fay, D.B. Ireland, J.P. Racamoto, W.D. Ross and A.M. Waxman, Field evaluations of dual-band fusion for color night vision, *SPIE Conference on Enhanced and Synthetic Vision 1999*, Verly, J.G. Ed., pp. 168–175 (1999).
  44. J. Sinai, J.S. McCarley and W.K. Krebs, Scene recognition with infra-red, low-light and sensor fused imagery, *Proceedings of the IRIS Specialty Groups on Passive Sensors*, pp. 1–9 (1999).
  45. W.J. Carper, T.M. Lillesand and R.W. Kiefer, The use of intensity-hue-saturation transformations for merging spot panchromatic and multispectral image data, *Photogram. Eng. Remote Sens.*, **56**, pp. 459–467 (1990).
  46. M.E. Ulug and L. Claire A quantitative metric for comparison of night vision fusion algorithms, *Sensor Fusion: Architectures, Algorithms and Applications IV*, Dasarathy, B.V. Ed., pp. 80–88 (2000).
  47. C.S. Xydeas and V.S. Petrovic, Objective pixel-level image fusion performance measure, *Sensor Fusion: Architectures, Algorithms and Applications IV*, Dasarathy, B.V. Ed., pp. 89–98 (2000).



# **Military Applications of Remote Sensing**

This page is intentionally left blank

## CHAPTER 10

# AUTOMATING THE ESTIMATION OF VARIOUS METEOROLOGICAL PARAMETERS USING SATELLITE DATA AND MACHINE LEARNING TECHNIQUES

R.L. Bankert, M. Hadjimichael, A.P. Kuciauskas, K.A. Richardson, F.J. Turk,  
and J.D. Hawkins

*Naval Research Laboratory*  
7 Grace Hopper Ave., Monterey, CA 93943-5502  
E-mail: [bankert@nrlmry.navy.mil](mailto:bankert@nrlmry.navy.mil)

Automated algorithms are being developed to assist U.S. Navy operational weather assessment and forecasting. Using supervised machine learning techniques, patterns and relationships are discovered in various satellite and meteorological data from which relevant classification and parameter estimation algorithms can be developed. Three applications of these techniques are discussed. A Geostationary Operational Environmental Satellite (GOES) image cloud type classifier is developed using expert-labeled data, specific image characteristic features and a 1-nearest neighbor classification routine. A tropical cyclone intensity estimation algorithm is developed using brightness temperatures and derived features of Special Sensor Microwave Imager (SSM/I) data, best-track intensity (ground truth) and a  $K$ -nearest neighbor routine. Knowledge Discovery from Databases (KDD) methodology is employed to develop algorithms to estimate cloud ceiling height at remote locations. Developed over a two-year period, the database consists of hourly location-specific records of satellite data, numerical weather prediction data, and ground truth (METAR) cloud ceiling height observations. Data mining techniques are applied to produce cloud ceiling height estimation algorithms. All of the algorithms mentioned above exist at various stages of development and each has shown promising potential for operational use.

### 1. Introduction

Automated satellite interpretation algorithms and diagnostic tools that provide meteorological parameter estimation are being developed to assist the U.S. Navy weather observer and forecaster. Having reliable tools that can quickly analyze data from satellite, radar, numerical weather prediction models, ground observations, etc and provide specific information about the weather conditions of interest at a remote location would be very useful for any operational meteorologist. The algorithms discussed in this chapter are a small subset of the possible tools that could be created and focus on satellite data applications. Algorithms are developed using specific data patterns and relationships, discovered by pattern recognition and supervised learning techniques.

The advances in data collection made in recent years have indicated a need for a wide selection of sophisticated methods of data reduction and analysis. At the same time, significant increases in computer storage and computing power have made new artificial intelligence methods and procedures feasible. These methods are collectively referred to as data mining tools and include both traditional statistical methods and artificial intelligence machine learning algorithms. The work presented here relies on this set of tools.

Geostationary Operational Environmental Satellite (GOES) data is used to develop a cloud classification algorithm that can be used for both land and sea scene classifications. The algorithm takes advantage of all five GOES spectral channels (for a daytime classifier) and can be used for both GOES-8 (East) and GOES-10 (West) imagery. Training sample characteristic feature vectors are computed from the channel data. A nearest-neighbor routine is used to compute the similarity of the characteristic

feature vector of a testing sample with those of the samples in the training set [1].

Analysis of remotely-sensed data has become an increasingly valuable tool for determining tropical cyclone location and structure, formulating tropical cyclone intensity estimates, and predicting strength and movement. Aircraft reconnaissance missions stopped in 1987 in the western Pacific, leaving the Atlantic basin as the only region where tropical cyclones are routinely monitored by aircraft. As upper-level (nonprecipitating) clouds are essentially transparent within the passive microwave imagery [2], the low-level structure and circulation of tropical cyclones can, in most cases, be detected in Special Sensor Microwave/Imager (SSM/I) data. A  $K$ -nearest neighbor, where  $K \geq 1$ , classification algorithm is developed to estimate the intensity of a tropical cyclone using 85-GHz channel and derived rain rate features. The algorithm is designed to work on tropical cyclones in any ocean basin. As opposed to the single nearest-neighbor used in the GOES cloud classifier, multiple training set neighbors can be used by the  $K$ -nearest neighbor algorithm to estimate the intensity of any tropical cyclone (as seen in the SSM/I imagery).

Knowledge Discovery from Databases (KDD) methods are used to study the feasibility of estimating cloud ceiling height at remote locations where ground observations are not available. Data mining tools are applied to a database of environmental records to uncover patterns that represent physical laws implicit in the data. The goal is to find relationships in satellite and numerical weather prediction data that can provide estimates of cloud ceiling height, an important aviation parameter. Over a two-year period, hourly records of data from various geostationary and polar-orbiting satellites and output from the Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS™) at a total of 45 specific locations in three geographic regions (U.S. West Coast, Adriatic Sea, and Korean Peninsula) were collected and processed. Data relationships that estimate cloud ceiling height were uncovered through data mining processes and the resulting algorithms tested. Precipitation rate and accumulation will be studied and discussed in future papers.

## 2. GOES Cloud Classification

Automated scene classification of satellite imagery is well researched. Meteorological applications include classification of surface and cloud types in polar regions [3, 4, 5], the discrimination of ice and water clouds [6], the separation of clouds and snow [7, 8], the classification of ocean clouds [9], and the discrimination between single- and multi-layered clouds [10].

Using the research and development of an Advanced Very High Resolution Radiometer (AVHRR) cloud classifier [11] as a guide, a cloud type classification algorithm is developed for GOES daytime imagery. Similar to the AVHRR classifier, supervised learning and pattern recognition methodologies are used to establish a training data set, characteristic feature vector, and classification technique. Since nonparametric and nonlinear classification methods such as neural networks and  $K$ -nearest neighbor provide superior classification analysis [12] as compared to algorithms such as linear discriminant analysis,  $K$ -nearest neighbor is chosen as the classification algorithm. In this application,  $K=1$ .

Taken from a time period of February, 1999 through August, 2000, a training set of expert-labeled 16x16 km samples (Figure 1) is created from GOES-8 and GOES-10 data for the daytime classifier. These training samples were independently classified by three satellite interpretation experts. The class types from which the experts had to choose are listed in Table 1. No mixed cloud-type classes are used. Only those samples given the same classification by all three experts were saved in the training set. The daytime

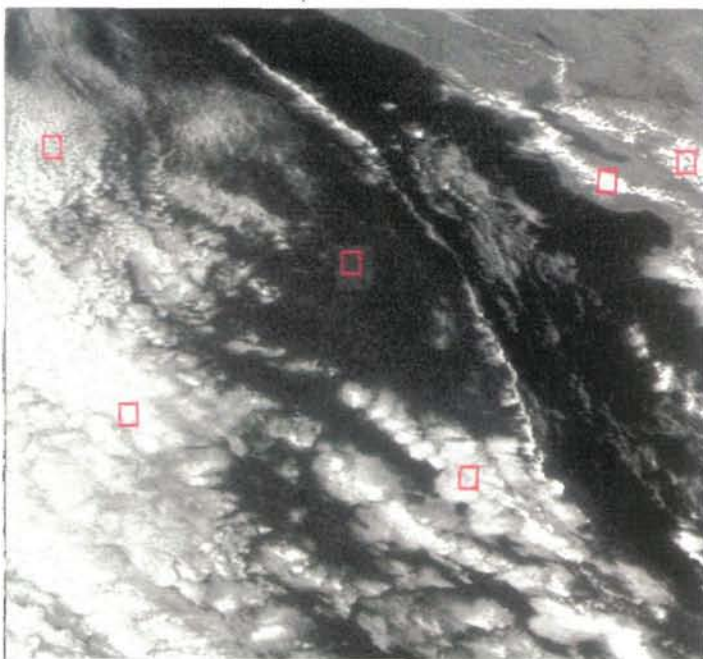


Figure 1. GOES-10 image (central California coast) with red boxes indicating example 16x16 km boxes used as training samples in GOES cloud classifier.

training data was separated into two sets - land (5313 samples) and water (5937 samples). Since the visible channel data is not valid at night, training data for a nighttime classifier will be developed separately using data from the infrared channels.

GOES imagery consists of five spectral channels (in microns): 0.55-0.75 (channel 1 - visible), 3.8-4.0 (channel 2), 6.5-7.0 (channel 3), 10.2-11.2 (channel 4), and 11.5-12.5 (channel 5). Data from channel 1, the visible channel, has a resolution of 1 km at nadir.

Table 1. Class types assigned for GOES cloud classification training data set.

Stratus (St)	Cumulus Congestus (CuC)
Stratocumulus (Sc)	Cs associated with deep convection (CsAn)
Cumulus (Cu)	Cumulonimbus (Cb)
Alto cumulus (Ac)	Clear skies (CI)
Altostratus (As)	Haze, sand, smoke, or dust (Hz)
Cirrus (Ci)	Ground snow (Sn)
Cirrocumulus (Cc)	Sunglint (Sg)
Cirrostratus (Cs)	

Channels 2, 4, and 5 are at 4 km resolution, and channel 3 has a 4 km x 8 km resolution. Using data from all five channels, over 100 characteristic features are computed for each training sample. These features include maximum, minimum, mean, median, mode, and standard deviation within the sample for each channel and selected channel differencing images (e.g. channel 2 - channel 4). The higher resolution of the visible channel allows

textures (contrast, homogeneity, entropy, etc) to be computed for that channel [13]. Textures provide a representation of the spatial distribution of the data within the sample. Latitude, longitude, time-of-year, and climatological sea surface temperature (water samples only) are also added as features for each sample.

In addition to increasing classifier processing time, using a large number of features degrades classifier performance due to the introduction of redundant and irrelevant information. For these reasons, a feature selection algorithm is applied to the training data to determine a subset of features which optimizes classification accuracy. In general, there are three components to a feature selection routine: the search algorithm, the evaluation function, and the classifier [14, 15]. The classifier uses the feature subset, found by the search algorithm, that maximizes the evaluation function. The feature selection algorithm employed here is a variation of the backward sequential selection (BSS) algorithm<sup>1</sup> and a 1-nearest neighbor evaluation function. Since the classifier and evaluation function are the same, the problem of bias in either algorithm is avoided. Using multiple feature subsets (as defined by the BSS algorithm), leave-one-out cross validation tests are performed on the training data set. Leave-one-out cross validation involves testing on each individual sample while using the remaining samples as the

*Table 2. GOES Classifier Selected Features (Daytime).*

<u>LAND</u>	<u>WATER</u>
Satellite (GOES-8, 10)	Satellite (GOES-8, 10)
Latitude	Latitude
Time-of-year	Time-of-year
Channel 2 minimum	Channel 1 median
Channel 2 median	Channel 1 standard dev
Channel 3 median	Channel 2 maximum
Channel 4 maximum	Channel 2 mean
Channel 4 range	Channel 3 maximum
Channel 4 median	Channel 4 minimum
Channel 5 minimum	Channel 4 range
Channel 5 median	Channel 4 mean
Channel 1 std dev texture (mean of 4x4 km boxes)	(Channel 4 - Channel 5) mean
Channel 1 entropy	Channel 1 mean difference texture
Channel 1 mean sum texture	Channel 1 entropy
Channel 1 mean sum texture (std dev of 4x4 km boxes)	Sea surface temperature (climo)

<sup>1</sup> The BSS algorithm is a process in which features are progressively removed from the feature subset until performance does not improve.

training set. The feature subset that produces the highest leave-one-out cross validation classification accuracy is the final selected set. Table 2 is a list of the selected features used in the GOES daytime classifier. With the selected feature subset, leave-one-out cross validation testing on both land and water training sets produced approximately 90% overall accuracy.

When an unclassified (testing or operational) 16x16 km GOES image sample is presented to the classifier, the similarity distance (equation 1) is computed between that sample and each training set sample.

$$\sum_i (\text{testing feature}_i - \text{training feature}_i)^2 \quad (1)$$

Each feature value used in (1) has been normalized, and the summation is over all features. The class type of the training sample with the minimum similarity distance is assigned to each pixel in the 16x16 km testing sample (at the visible channel 1 km resolution). See the 1-nearest neighbor illustration in Figure 2.

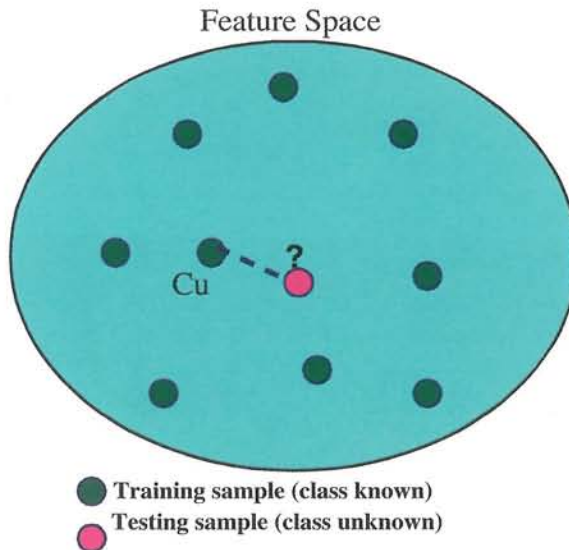


Figure 2. 1-nearest neighbor illustration. Testing sample (“?”) is assigned to be Cu since the nearest training sample in feature space is a Cu type.

## 2.1 Real-Time Image Processing

For a complete GOES image classification, the 16x16 km classification boxes overlap each other such that each individual pixel is classified by the algorithm four times (excluding those pixels on image edges). The final classification of an individual pixel is determined by simple majority with ties broken by random selection. In addition, the training set choice (land or water) for classification boxes that fall over a coastline is determined by whichever type (land or water) covers the most pixels in that box.

To present a more accurate classification of an image on the pixel level, the visible



channel albedo (corrected for the solar zenith angle) for each pixel is checked. If that value is less than  $12\%^2$ , then the pixel is re-classified as clear (Cl), regardless of the

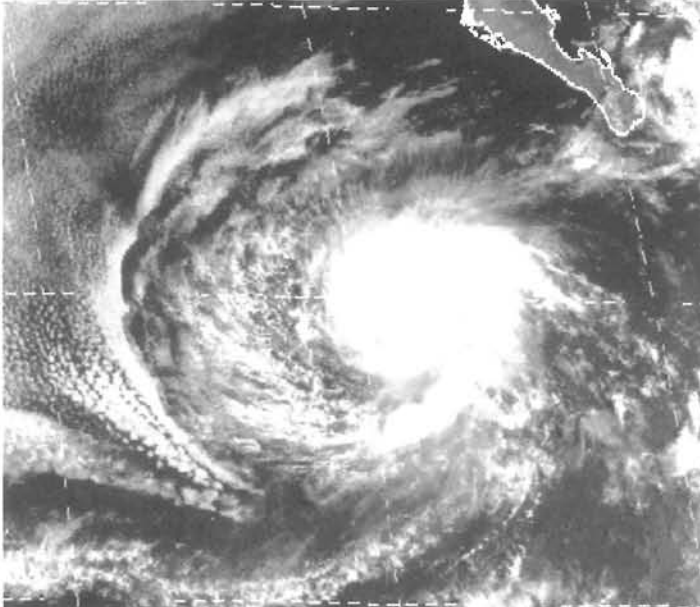


Figure 3a. GOES-10 visible image of Hurricane Flossie (28 Aug 2001, 1700 UTC) in the Eastern Pacific Ocean.

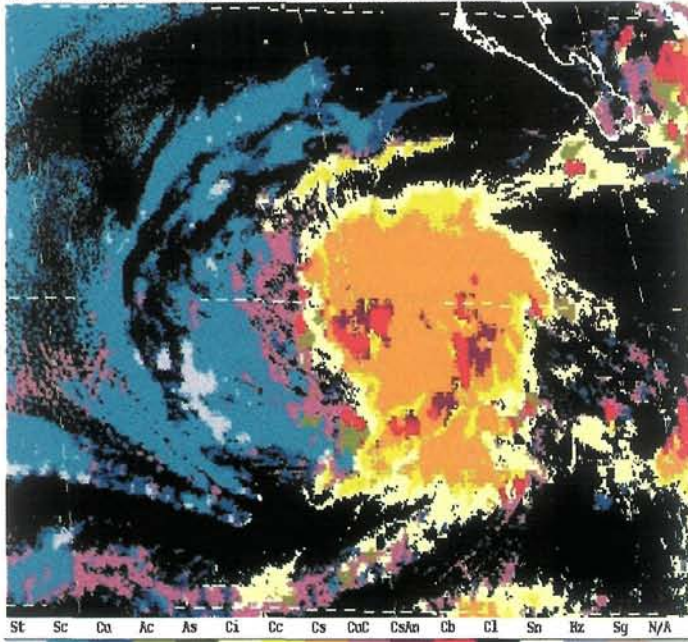


Figure 3b. Cloud classification image of Figure 3a.

- 2 The albedo threshold of  $12\%$  was chosen as a conservatively low value for which the non-clear classes would not fall below.



initial classification. Additionally, due to observed classifier problems with ground snow misclassifications and known satellite-sensed properties of snow [7], pixels initially classified as snow are checked, using the reflectance portion of channel 2 and the visible channel. If the channel 2 reflectance is greater than 3%, the pixel is re-classified as stratocumulus (Sc) or if the channel 1 albedo is less than 30%, the pixel is re-classified as clear.

Classification images are created using color-coded representations of the various classes. See Figures 3a and 3b for an example classification image.

Known limitations of the GOES daytime cloud classifier include the inability of the system, in its present form, to classify mixed cloud samples. Every 16x16 km box is classified with one of the "pure" classes even if it is a truly mixed sample. Situations such as thin cirrus (Ci) over low clouds (St or Sc) have been noted to be classified as As or Ac. The classifier is also limited by the amount and variability of the training data in each class. Representing the entire universal set of these classes as they are distributed over time (seasons) and space (latitude and longitude) is virtually impossible. A web page of real-time GOES daytime cloud classification output can be found at

<http://www.nrlmry.navy.mil/sat-bin/clouds.cgi>

## **2.2 Future Plans**

A GOES nighttime cloud classifier is currently in development. Similar to the daytime classifier, expert-labeled samples are being collected to provide land and water training sets with features to be computed from four of the five spectral channels (no visible channel data). Alternative feature selection and classification algorithms will also be studied. Other plans include examining the use of a surface albedo look-up table for pixel postprocessing (replacing 12% threshold) and developing automated classification algorithms for other geographical regions. Using data from the next generation of geostationary satellites would be the logical choice to accomplish this goal. These satellites include the European Meteosat Second Generation (MSG-1, successful August 2002 launch) and the Japanese Multifunctional Transport Satellite (MTSAT-1R).

## **3. SSM/I Tropical Cyclone Intensity Estimation**

The reliable Dvorak method [16, 17] provides manually estimated tropical cyclone intensity from visible and infrared satellite imagery using subjective pattern recognition and applicable rules. An objective Dvorak technique (ODT) [18] was developed to eliminate much of the subjectivity of the Dvorak method.

Using SSM/I images to examine tropical cyclone structure has an advantage when compared with the limitations of other imagery types [19]. Rainbands and a tropical cyclone center (when it exists) can be seen in the 85-GHz channel images when this structure is obscured by upper-level clouds as seen in visible and infrared imagery. Figures 4a-c illustrate this contrast in image types as the structure of Typhoon Joan is obscured in the visible and infrared imagery, but is apparent in the SSM/I 85-GHz image. Extracting this tropical cyclone structural information will be invaluable in developing an automated algorithm to estimate tropical cyclone intensity.

One attempt to automate tropical cyclone intensity with SSM/I data used empirical orthogonal functions as inputs to a neural network [20] and demonstrated the potential of using SSM/I data in an automated system. The research described here examines the viability of using SSM/I image characteristic features and local information to estimate tropical cyclone intensity (as measured by the maximum wind speed). Similar to the

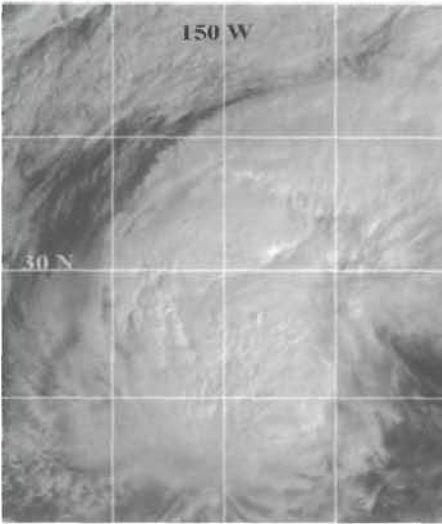


Figure 4a. GOES-9 visible image of Typhoon Joan (2232 UTC 22 Oct 1997).

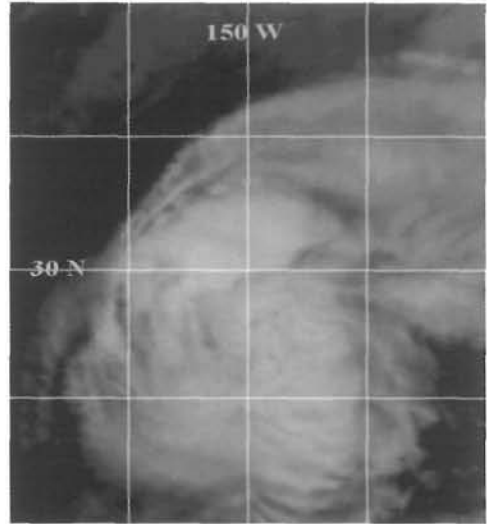


Figure 4b. Infrared image of Typhoon Joan.

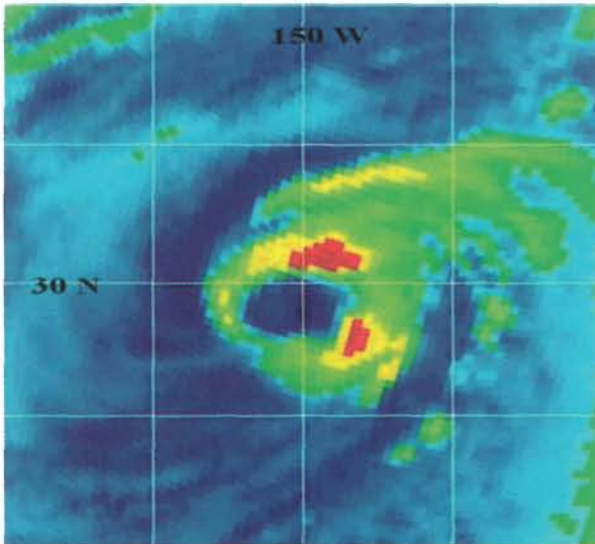


Figure 4c. SSM/I 85-GHz image of Typhoon Joan. The red areas indicate low brightness temperatures (high scattering due to large ice particles - convective precipitation) and blue areas indicate higher brightness temperatures (no or limited convection).

GOES cloud classification algorithm, a set of SSM/I tropical cyclone training data is needed. Using the available data and information, including Dvorak estimates, surface observations, and reconnaissance reports (Atlantic and eastern Pacific), a best-track maximum wind speed has been determined for each SSM/I training sample by either the Joint Typhoon Warning Center (JTWC) or the National Hurricane Center (NHC). This best-track intensity will serve as the ground truth for this application development and has an estimated error range of +/- 15 kts.

The SSM/I instrument uses a suite of passive microwave channels: 19, 22, 37, and 85 GHz. The 85-GHz channel has an advantage over the other microwave channels with its ability to detect convective precipitation due to high scattering from ice particles. The 85-GHz image data have a spatial resolution of 13 km x 15 km. Applying an interpolation algorithm [21], this data can be mapped to 1-2 km resolution. In addition to the 85-GHz data, rain rate “image” data is also used to extract feature characteristics that will be presented to a  $K$ -nearest neighbor algorithm. The rain rate “image” is a derived product from multiple SSM/I channels [22]. Example 85-GHz and rain rate images are shown in Figure 5.

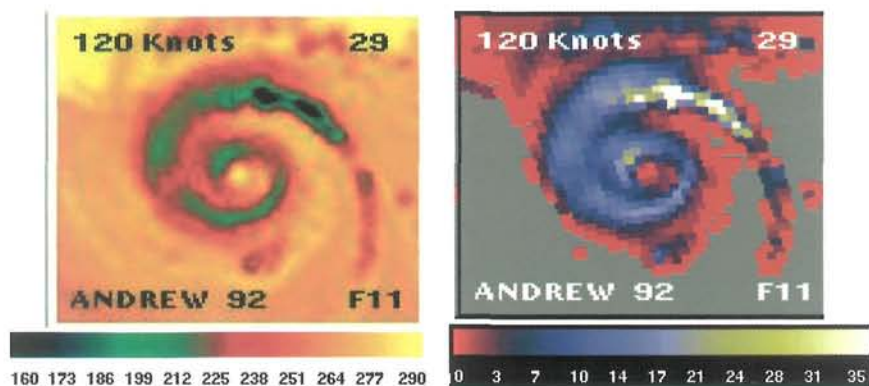


Figure 5. 512x512 SSM/I 85 GHz image (left, in degrees K) and 512x512 SSM/I-derived rain rate image (right, in mm/hr) for Hurricane Andrew (2252 UTC 25 Aug 1992).

### 3.1 SSM/I Data and Features

SSM/I data from 175 tropical cyclones that comprise the period of 1988-2001 and all relevant ocean basins are used to train and test the tropical cyclone intensity estimation algorithm. Features are computed from 1297 SSM/I images that are 512x512 (pixels at approximately 1 km resolution), centered on the tropical cyclone center, and have an associated best-track intensity. Over 100 characteristic features are computed from the 85-GHz and rain rate data for each image. These features include spectral characteristics which are simple statistical measurements (maximum, minimum, mean, median, mode, range of values, and standard deviation) of both 85 GHz and rain rate imagery. Textural feature values of the 85 GHz data are also computed and include contrast, entropy, angular second moment, and local homogeneity [13]. As noted for the GOES cloud classifier, they provide a representation of the spatial distribution of the brightness temperatures within the image.

Both the entire image and the inner 150x150 pixel region are divided into quadrants to extract information, in a general sense, about the convective organization in the tropical cyclone environment. Summations of all pixel values within each quadrant (both 85-GHz and rain rate) are saved as features. Features that are more tropical cyclone specific include: enclosed eye (yes, no), size of enclosed eye, and banding and concentric ring features. Latitude, longitude, and date (relative to peak activity) are also features [23]. A summary of all features can be found in Table 3.

Table 3. Summary list of extracted image (85 GHz and rain rate) features.

<b>General</b>
Spectral
Textures
Pixel value summation (complete image, quadrants, inner region)
<b>Local</b>
Latitude
Longitude
ABS(Yearday - 237)
Sea surface temperature (climatological)
<b>Segmented Image</b>
Pixel count
E-W, N-S extent
<b>Tropical Cyclone Specific</b>
Enclosed eye (yes/no)
Size of enclosed eye
Max/min/range of thresholds for enclosed eyes
Warmest pixel value in tropical cyclone center (WPC)
Representative surrounding temperature (ST)
WPC - ST
Banding and concentric ring features

### 3.2 Feature Selection and *K*-Nearest Neighbor

Similar to the manner in which it was applied to the GOES cloud classification training data, a feature selection algorithm is applied to the SSM/I tropical cyclone data. The notable difference here is that the evaluation function and classifier is a *K*-nearest neighbor algorithm, where  $K \geq 1$ .

Using the single nearest-neighbor distance (see Figure 2) as the standard for inclusion, those training samples within a distance factor (1.75 X nearest-neighbor distance) are used to estimate the testing sample intensity. A simple averaging technique is performed on those *K*-nearest neighbor best track intensities. A leave-one-out cross validation test is applied to a selected training set (942 of the 1297 samples) and the intensity root mean square error (RMSE) is computed relative to each feature subset. The error for any given training sample is the difference between the *K*-nearest neighbor estimated intensity and the best-track intensity. The reduced subset that produces the minimum RMSE after the search and *K*-nearest neighbor evaluation is the final selected set. These selected features are listed in Table 4.

### 3.3 Testing Results

The final training set increased by 237 samples (with the addition of tropical cyclones from 1999-2001), after completion of the feature selection, for a total of 1199 training samples. The testing set consists of 98 samples from four tropical cyclones:

Table 4. SSM/I tropical cyclone selected feature list.

Latitude
Longitude
ABS(Yearday - 237)
85-GHz - Pixel summation - NE quadrant (512x512)
85-GHz - Pixel summation - NE quadrant (inner 150x150)
85-GHz - Segmented - normalized mean radius
85-GHz - Range of pixel values (512x512)
85-GHz - Range of temperature thresholds for which enclosed eye exists
Rain rate - Average pixel value for those > 0 mm/hr (512x512)
Rain rate - Number of pixels > 0 mm/hr (inner 150x150)
Rain rate - Pixel summation - SE quadrant (512x512)
Rain rate - Pixel summation - NW quadrant (inner 150x150)
Rain rate - Mean pixel value (512x512)
Rain rate - Mode pixel value (512x512)
Rain rate - Banding feature - maximum summation of pixels on concentric rings (> 3 mm/hr)

Hurricane Guillermo (August, 1997), Supertyphoon Oliwa (September, 1997), Supertyphoon Paka (December, 1997), and Supertyphoon Babs (October, 1998).

Using the 15 selected features (Table 4) as the representative vector for all of the training and testing samples, the testing set intensities are estimated within the  $K$ -nearest neighbor algorithm and compared to the best-track ground truth intensity estimates. For the 98 testing samples, the overall RMSE is 19.0 kts with an average absolute error (AAE) of 15.3 kts, and an average percentage error (APE) of 25.0%. A majority of the samples (56%) had an intensity estimate within 15 kts of the best-track intensity, but 7% had an error greater than 30 kts. See Table 5 for a summary of the errors statistics for each of the four tropical cyclones in the testing set.

Many of the high intensity estimation errors can be found with images associated with Supertyphoon Oliwa. As an example, an 85-GHz image of Oliwa in the early stages of development (best-track intensity of 35 kts) can be seen in Figure 6. The tropical cyclone intensity estimation algorithm is apparently doing a poor job of interpreting the convective area around the center of the cyclone. The convective banding (or lack thereof) is not being adequately handled by the algorithm. The algorithm produced an intensity estimate of 99.1 kts (64.1 kt error) for this image.

Figure 7 is a time series plot of the best-track intensity and  $K$ -nearest neighbor intensity estimation for Supertyphoon Paka. All four tropical cyclones had  $K$ -nearest neighbor time series plots that exhibited "spikes" during the lifetime of the cyclone. This high variability in the SSM/I intensity estimate is inconsistent with the best-track data and contributes to the computed error. However, some of these spikes could be representing actual strengthening or weakening of the cyclone that is not captured in the smooth best-track data.

### 3.4 Discussion

The snapshot approach (no historical information) used in this first version of the SSM/I tropical cyclone intensity estimation algorithm can be improved upon when adding the history of past intensity to the set of features [23]. This is one experiment that could be performed in future work on the recently expanded data set. Additionally,



Table 5. *K*-NN testing results (98 total samples). RMSE - Root Mean Square Error; AAE - Average Absolute Error; APE - Average Percent Error.

Tropical Cyclone	RMSE (kts)	AAE (kts)	APE (kts)
Oliwa (25 samples)	25.9	20.1	39.3
Guillermo (24 samples)	19.3	17.3	25.8
Paka (27 samples)	14.1	11.7	17.1
Babs (22 samples)	14.3	12.2	17.5
OVERALL	19.0	15.3	25.0

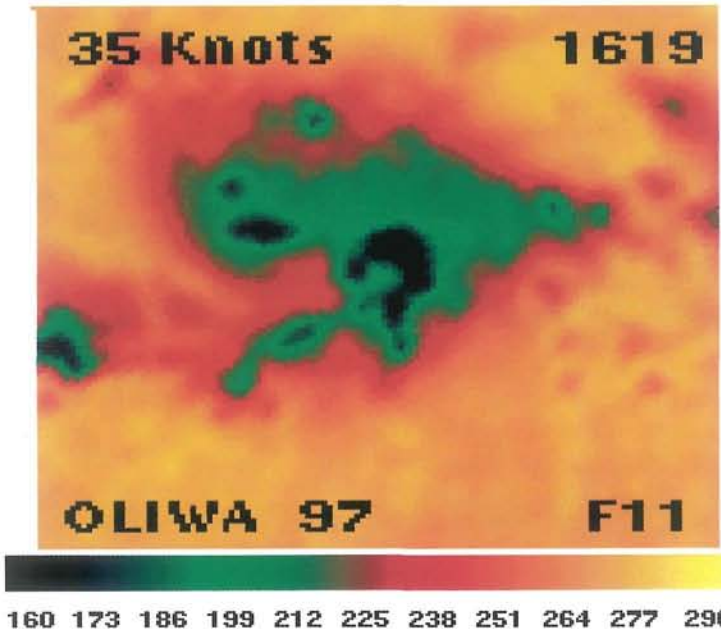


Figure 6. A 512x512 85-GHz image centered on tropical cyclone Oliwa (2007 UTC 6 Sep 1997) at a time of relatively low intensity with extensive areas of cold brightness temperatures (blue and green areas) around the tropical cyclone center.

if the larger training data set were used in the feature selection algorithm, a more optimum feature subset may have resulted. This possibility is worthy of investigation.

While improvement in the methodology described here is needed to provide accurate intensity estimates in an operational environment, it is important to note that development and evaluation of the algorithm is dependent upon the accuracy of the best-track intensity of the training and testing samples. These intensities are estimated to have an error range of +/- 15 kts.

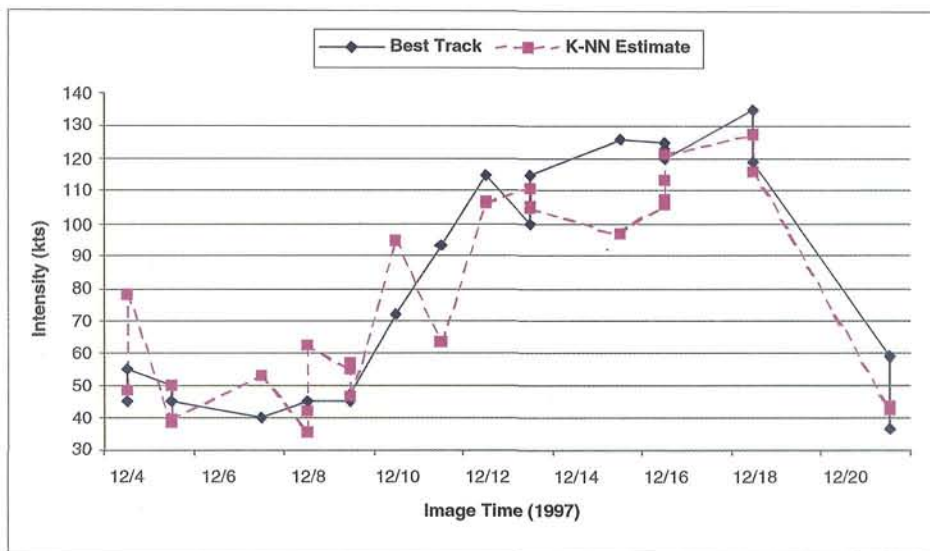


Figure 7. Time series plot of best track intensity and K-nearest neighbor intensity estimate for Supertyphoon Paka.

Given the wide variety of possible imagery characteristics, continuing to increase the number of unique training samples will improve the estimation algorithm. Examining the use of feature characteristics from other SSM/I channels, using variations of  $K$ -nearest neighbor estimation, and studying other methods to use tropical cyclone history or other ancillary information may produce alternative procedures to improve the algorithm.

Beyond experimentation with the current algorithm, investigating the use of other data sources and the use of output from other methodologies (e.g., Advanced Microwave Sounding Unit (AMSU) data and Advanced Objective Dvorak Technique (AODT) output) is anticipated. Additionally, other pattern recognition and artificial intelligence technologies (e.g., fuzzy rule-based system) should be applied to compare performance and find the best approach for estimating tropical cyclone intensity using SSM/I data.

#### 4. KDD - Cloud Ceiling Height Estimation

Currently, the primary approach for correcting numerical weather prediction model output is via model output statistics (MOS) [24, 25]. MOS is a technique for postprocessing numerical weather prediction output. MOS correlates model output and other climatic variables with surface observations using multiple linear regression, in order to predict observed variables such as cloud base height [25, 26, 27, 28]. Several drawbacks can be identified with using only MOS for model correction:

- 1) The latest surface observation predictors are often the most important terms in MOS forecast equations and dominate over all model-based terms [24, 25]. In the application described in this section, situations when no local observations are available is of primary interest.
- 2) MOS methodology is generally limited to multiple linear regressions. Thus, any nonlinear solutions will be missed. In particular, any MOS result will be limited

by the statistical model imposed on the solution.

The field of knowledge discovery from databases, KDD, refers to the overall process of discovering useful information from data, including issues of storage and access, all forms of data preparation, addition of necessary domain knowledge, and interpretation and visualization of the results. Data mining refers to the specific algorithms applied to the data in order to extract patterns [29]. Data mining techniques include both traditional statistical methods and artificial intelligence machine learning algorithms. Some of the commonly used data methods include inductive machine learning, regression, clustering, summarization, generalization, dependency modeling, and link analysis [30, 31]. KDD has been successfully applied in both business and scientific fields. The mining of scientific data is most relevant to this application, and is well reviewed in [32]. Some notable systems and results include SKICAT [33], JARtool [34], and OASIS [35]. SKICAT (Sky Imaging Cataloging and Analysis Tool) uses decision tree methods to predict the classes of faint astronomical objects in photographic image data. JARtool (JPL Adaptive Recognition Tool) learned to recognize small volcanoes in satellite images of the surface of Venus. The OASIS data mining environment was designed for discovery and visualization in large geophysical datasets. It has been applied successfully in the study of spatio-temporal features of cyclonic storms. The project described in this section applies the complete KDD process to the problem of determining remote meteorological parameters.

#### **4.1 Meteorological Application Background**

U.S. Navy weather observing and forecasting operations would be greatly assisted with the immediate assessment of remote meteorological parameters when ground observations are not available. To this end, numerical weather prediction data and satellite data from various sensors and platforms are being used to develop automated algorithms to assist in operational weather assessment and forecasting. Data mining methods, used in a Knowledge Discovery from Databases (KDD) procedure, are applied to cloud ceiling height estimation at remote locations using appropriate geostationary and polar orbiting satellite data in conjunction with Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS™) data. Data mining methods have determined an algorithm to diagnose these sensible weather elements more accurately than numerical weather prediction or satellite methods alone. Further detail about the initial design of the study, data, methods, and comparisons to other methods can be found in [36]. Methodology overview and results from the data mining work are presented here.

#### **4.2 Data Sources**

In order to discover the relationships between a variety of physical variables, both calculated and measured, a database must be created from a “fusion” of data from various sources. A unique meteorological research tool consisting of a database of numerical weather prediction (COAMPS™) output, satellite data, climatology, and ground truth observations (METAR) has been created for use in data mining. COAMPS™ output parameters, coincident satellite parameters (including both geostationary and polar-orbiting data) and climatological information are extracted/computed at 45 METAR observation sites. Automated data collection routines have been written and data has been collected hourly for over two years (July, 2000



through September, 2002). Data mining techniques have been applied to study cloud ceiling height and rain accumulation diagnosis. The cloud ceiling height study will be discussed here.

#### **4.2.1 Numerical Model Data**

COAMPS<sup>TM</sup> is the numerical weather prediction model used to generate output values of selected relevant parameters. The model is run in three geographic regions, U.S. west coast, Adriatic Sea, and Korean peninsula, and configured with three nested grids (Figure 8) with 81 km, 27 km, and 9 km resolution from outermost to innermost grid. There are 33 grid levels in the vertical depth of the atmosphere. COAMPS<sup>TM</sup> is run for a 12-hour forecast cycle for each of these domain configurations at 00 UTC and 12 UTC every day. Interested readers can find additional COAMPS<sup>TM</sup> details in [37].

The closest land grid point (within each of the 9 km domains) to the 45 (18 West Coast, 14 Adriatic, 13 Korea) METAR stations is determined and COAMPS<sup>TM</sup> output values at those grid points for each hour are extracted and written to the database. Table 6 is a list of those COAMPS<sup>TM</sup> parameters.

In addition to extracting values for the database, COAMPS<sup>TM</sup> output can be viewed in static or animated 2D form over the appropriate domains for further interpretation and analysis.

#### **4.2.2 Satellite Data**

Data from three geostationary satellites, GOES-10, European Meteosat-7 (Meteorological Satellite), and the Japanese GMS-5 (Geostationary Meteorological Satellite) are extracted and added to the appropriate records in the database. This data will consist of all channel data at a given pixel whose center is closest to the latitude/longitude of each of the METAR stations. All visible channel data is corrected for the solar zenith angle. In addition to the channel data, a cloud optical depth algorithm [38] is applied and a GOES-only low cloud product [39] is derived, with their respective values extracted and stored.

National Oceanic and Atmospheric Administration (NOAA) polar-orbiting Advanced Very High Resolution Radiometer (AVHRR - local area coverage (LAC) and global area coverage (GAC)) data and Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave Imager (SSM/I) polar-orbiting data are also extracted and stored in the database. The AVHRR LAC and GOES data records include a derived cloud classification in addition to the channel data. In addition to the various microwave channel values, environmental data records (EDRs) are computed from the SSM/I channel data. These parameters include rain rate, cloud liquid water, and precipitable water.

Using appropriate COAMPS<sup>TM</sup> and satellite data, a cloud top height value is derived for all sensors except the SSM/I. Using satellite-based algorithms [40], rain rate and accumulation values are computed from a combination of geostationary infrared and passive microwave satellite data. Similar to the COAMPS<sup>TM</sup> output, satellite imagery can be viewed in static or animated 2D form. In addition to this visualization, monitoring tools have been developed to allow for a quick view of model and satellite retrieval performance.

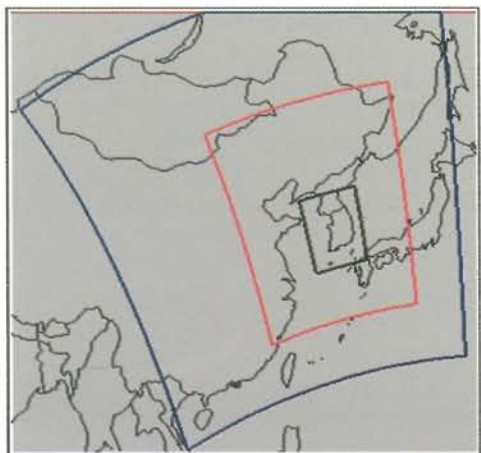
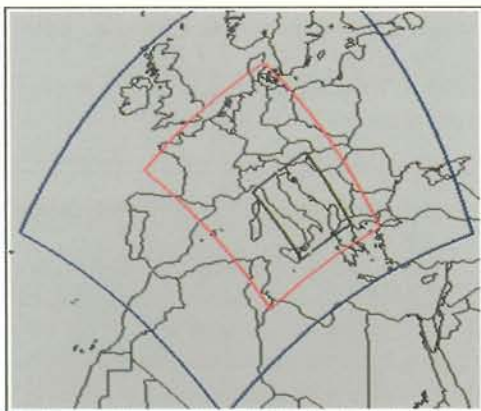
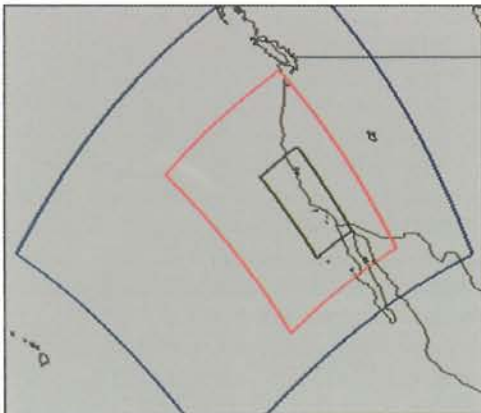


Figure 8. COAMPS nested grid locations for U.S. West Coast (top), Adriatic (middle), and Korea (bottom). Locations of interest (METAR stations) are inside the inner-most mesh.

#### 4.2.3 Ground Truth Data

All collected hourly METAR reports for the 45 selected stations are parsed, with sensible weather elements stored in the database. These weather parameters represent the ground truth and are the dependent variables in the subsequent search for patterns which relate satellite and model variables to locally observable parameters. Cloud ceiling height is the first parameter examined in the data mining portion of the KDD process. Rain rate and rain accumulation analysis will follow.

#### 4.3 Database Development

All parameters discussed in Section 4.2 are computed, processed, collected, and stored in a single database. The database tables are updated once a day after all COAMPS™ model runs and post-processing has been completed. Each table row represents the available information for a particular location at a specific hour.

The flow of data involves five steps:

1. Data generation/collection.
2. Data cleaning and pre-load processing.
3. Data loading in each individual, source-specific table.
4. Data post-load processing (calculate/update various derived fields).
5. Data consolidation: generating an *Event* record for each date/time where complete information is available (i.e., data from all three sources).

The database is organized in a star schema as shown in Figure 9. The key of each table is the day-time group and the location ID. The Climatology and Location tables are constant-valued reference tables, while the NWP (COAMPS™), OBS (METAR), and Satellite tables are updated daily with new data, consisting of records for each specific location and day-time group.

Some pre-load steps include:

- Time rounding: adjusting the time stamp of METAR reports and satellite points to the closest hour, to correspond with the model data.
- Satellite filtering: recognizing missing data.
- Satellite derived products: low clouds, cloud optical depth, cloud classifications, environmental data records, etc.
- METAR report processing: computation of vapor pressure, cloud/no cloud, variable wind directions, etc.
- METAR cleaning: removing duplicate, later corrected, or mislabeled reports.
- COAMPS<sup>TM</sup>/Satellite combination products: using satellite infrared temperatures together with COAMPS<sup>TM</sup> profiles to determine cloud top height.

The data mining tools used here require as input a denormalized (flat) table. In other words, rows representing the location of interest will be selected from all database tables containing the required information and joined together to form a single *Event* record of up to 90 variables. Each row will represent all available information for one day/time at one location.

Table 6. COAMPS parameters extracted for each of the 45 locations.

10m u-wind	Cloud top (qc) temperature
10m v-wind	Cloud base height (RH)
10m temperature	Sea level pressure
10m dewpoint	Topography height
10m potential temperature	LCL
PBL depth	CCL
Surface wind stress	Visibility (derived)
Total downward radiation	Ceiling height (derived)
Net radiation	Bulk Richardson number
10m relative humidity	Ground wetness
10m sensible heat flux	Surface albedo
10m latent heat flux	Surface mixing ratio
Ground temperature	Total heat flux
Total rain	z/L
u*	Max vert. velocity in PBL
t*	Max TKE in PBL
q*	10m, 1500m temp diff
Surface roughness	Precipitable water
10m, sfc temperature diff	Cloud coverage
10m, sfc mixing ratio diff	Max mixing ratio in PBL
Cloud base height (qc)	1000mb, 850mb thickness
Cloud top height (qc)	Cloud/No Cloud

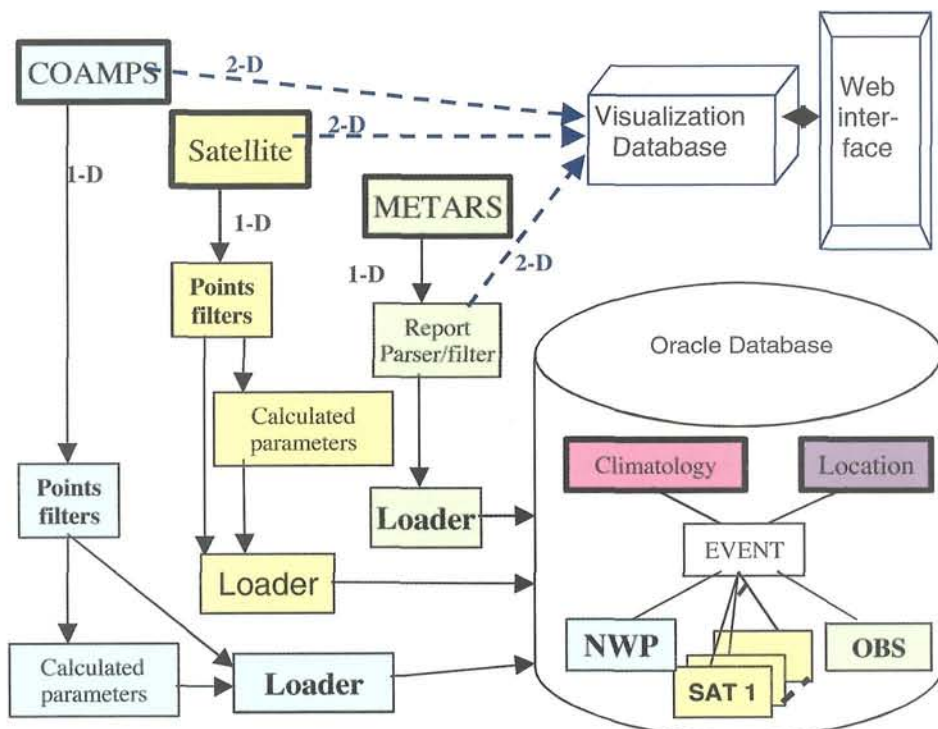


Figure 9. Data flow and schema. The three continuous data sources (COAMPS, Satellite, METARs), are stored for visualization through a web interface as two dimensional images, and are post-processed for storage in the database management system.

#### 4.4 Methodology

The primary tools selected for data mining are C5.0 and Cubist [41]. These were selected for their ease of use and well-recognized robustness. C5.0 generates decision trees, which are used for classification into categories. The Cubist program creates a set of rule-based predictive models, which are used for regression-type estimation of continuous values. Data was extracted from the database as a flat, ASCII text format file. Studies were done on each focus area independently, but combining each data from all locations within each focus area. Studies examining each individual station location independently rarely showed any improvement. This result is mostly likely because of the much smaller training set available for an individual location.

The data was randomly evenly divided into training and testing sets, although 10-fold cross validation was also used for error estimates. To achieve the best results, each experiment was decomposed into three components:

1. *Determination of cloud ceiling presence.* We used C5.0 to create a decision tree which could classify each record as "Cloud Ceiling" or "No Cloud Ceiling" to indicate cloud ceiling presence.
2. In locations with cloud ceiling presence, *determination of low cloud ceiling (< 1000m) versus high cloud ceiling.* Once again we used C5.0 to create a decision tree to classify the "Cloud ceiling presence" records.

Table 7. Classification and regression error comparisons.

West Coast Locations	Error (%)		Avg. error
			correlation
Method	Cloud presence	Low cloud detection	Cloud base height
COAMPS ceiling	25.5	50.2	213.0
			0.5
COAMPS variables	21.0	21.9	128
			.73
Satellite variables	12.2	22.2	170
			.52
<b>Fused variable set</b>	<b>11.1</b>	<b>19.9</b>	<b>126</b>
			<b>0.7</b>

3. In locations with low cloud, *determination of cloud ceiling height* is performed using a rule set generated by Cubist.

Learning experiments were based on three different sets of variables:

1. COAMPS<sup>TM</sup> variables only.
2. Satellite variables only (geostationary satellite for each region).
3. Fused (combined) COAMPS<sup>TM</sup> and Satellite variables.

#### 4.5 Results

The initial test has focused on cloud ceiling height for the U.S. West Coast locations during daylight hours. Overall, the data mining method outperformed all other methods. The data mining derived algorithm was compared to a COAMPS<sup>TM</sup> derived (using known physical relationships) cloud ceiling calculation, as well as the single source variable sets described above.

Figures 10 through 15 are plots of a sample location, KMRY (Monterey, CA METAR site), in several variations. Figure 10 is a comparison of KDD-diagnosed ceiling height for low clouds ("Predicted"), versus the true observed (METAR) ceiling height ("Ceiling\_Ht"). Figure 11 reflects the same data, but smoothed using a Bezier technique. The smoothed plots hide the individual variability, and expose a clear relationship between the cloud ceiling height determined through the KDD process and the ground truth values. The next two figures show the same plot, but with COAMPS<sup>TM</sup>-derived ceiling estimates included: both the COAMPS<sup>TM</sup>-derived ceiling, and a cloud base derived using the water mixing ratio. Note that, because COAMPS<sup>TM</sup> has difficulty determining low cloud ceiling cases, there are a great many cases plotted which are actually high cloud ceiling, or no ceiling captured in that data set. The smoothed plot shows this bias clearly. The final pair of figures plot only that data where both COAMPS<sup>TM</sup> and the KDD method indicate low ceiling. Once again, the smoothed plot shows a close correspondence between the KDD method and ground truth, and indicates

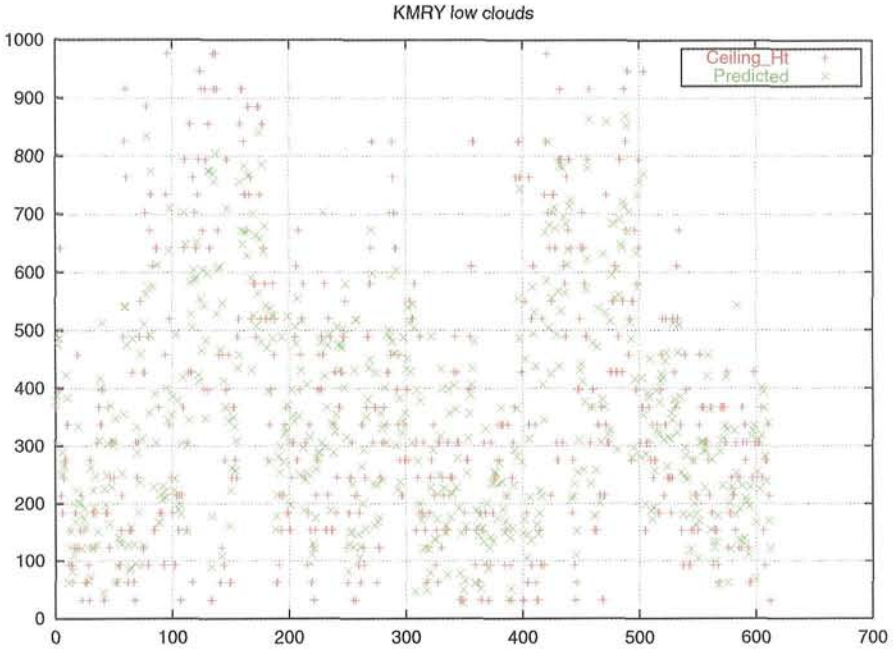


Figure 10. Plot of data mining-predicted low cloud ceiling heights ("Predicted") vs. ground truth ("Ceiling\_Ht"). In this and all following graphs, the x-axis denotes a observation record number, and the y-axis denotes the height in meters.

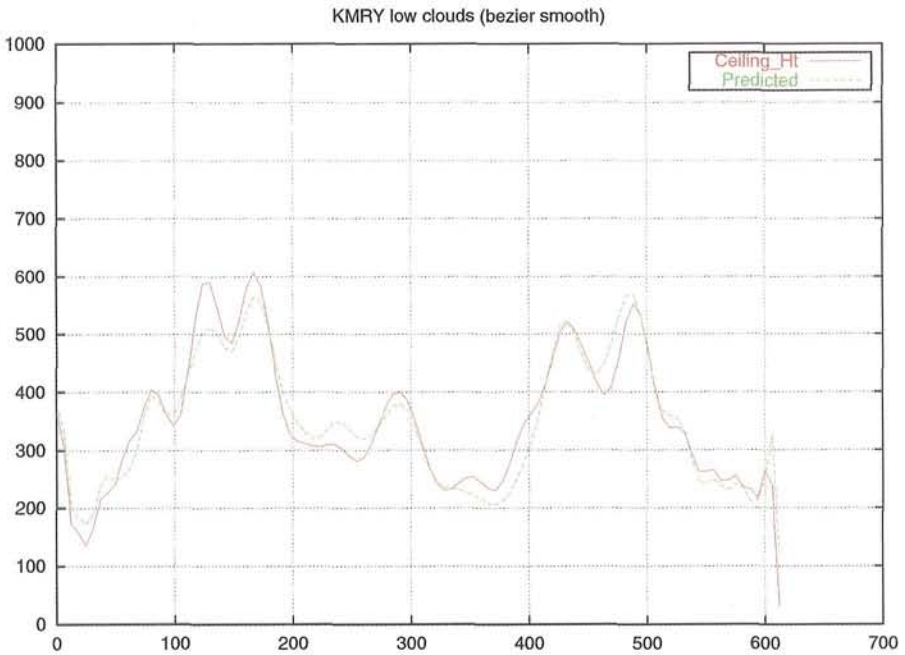


Figure 11. Same as previous figure, but with Bezier smoothing of datapoints.



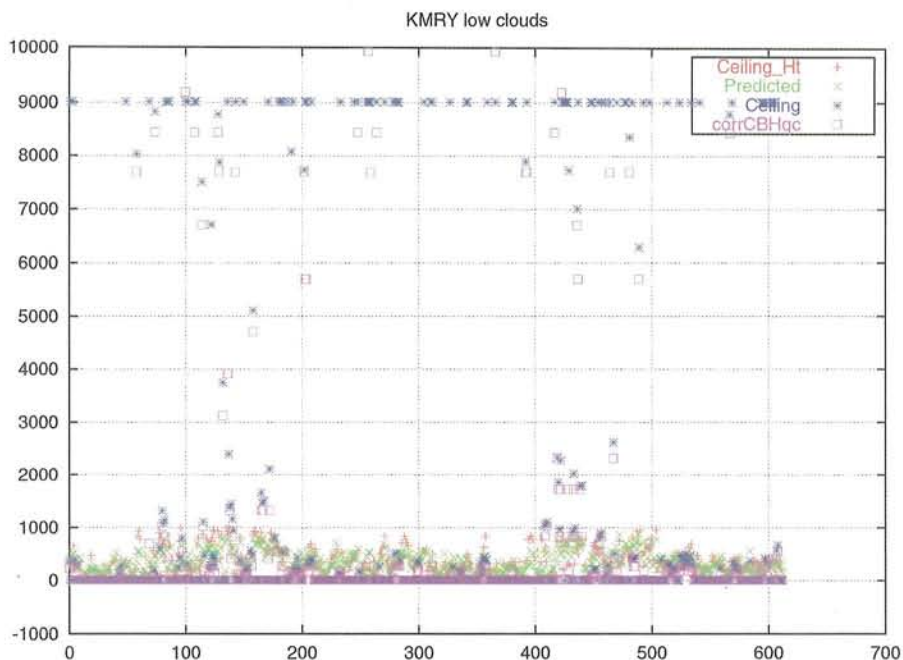


Figure 12. Plots of observed ceiling ("Ceiling\_Ht"), data mining predicted ceiling ("Predicted"), COAMPS derived ceiling parameter ("Ceiling"), and COAMPS water mixing ratio derived cloud base height ("corrCBHqc").

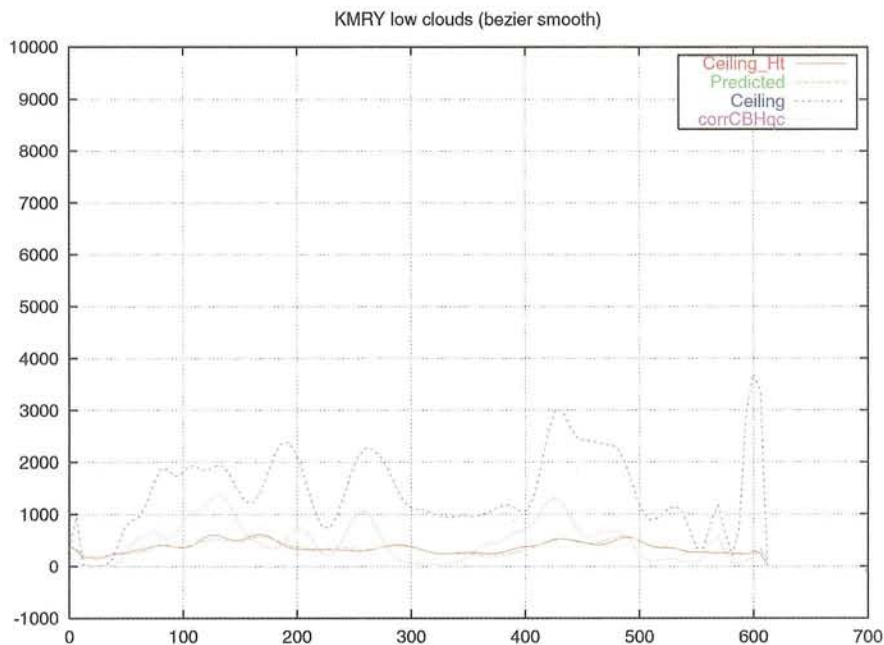


Figure 13. Same figure as previous, but with Bezier smoothing.

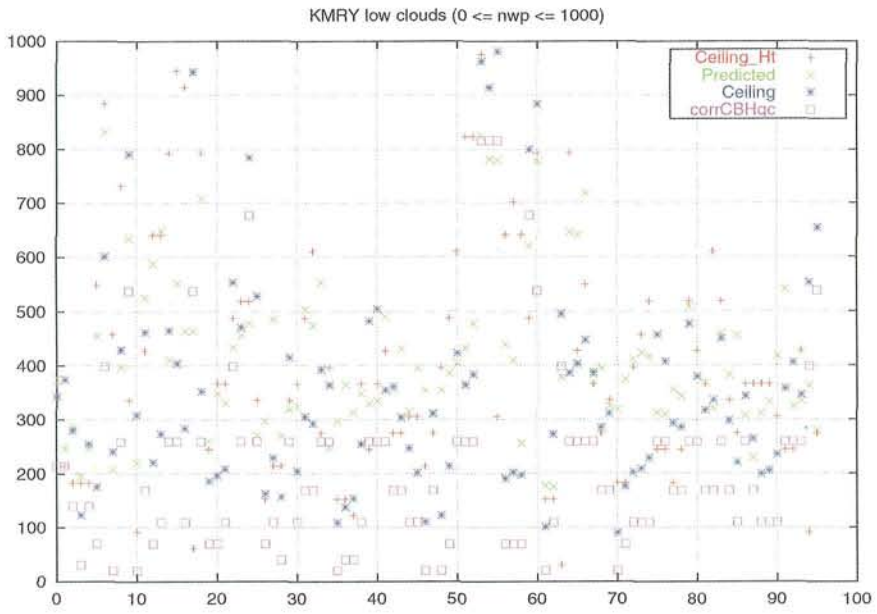


Figure 14. All plots, with only cases where COAMPS ceiling < 1000m..

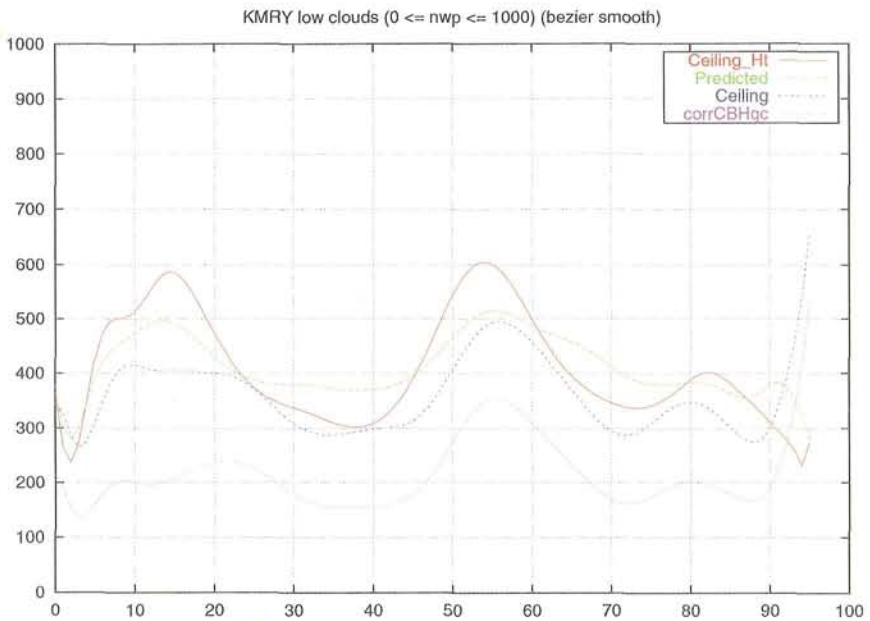


Figure 15. Same as previous plot, with Bezier smoothing..



a bias in the COAMPS<sup>TM</sup> estimates. Note that all the data was randomly shuffled and replotted, resulting in a similar correspondence in Bezier-smoothed plots, suggesting that the degree of correspondence is independent of any temporal information in the data.

Table 7 shows error comparisons for the entire west coast. The COAMPS<sup>TM</sup> ceiling method is the derived calculation produced by COAMPS<sup>TM</sup>. The remaining three methods are all created using C5.0/Cubist, on different sets of variables: COAMPS<sup>TM</sup> variables only (except the ceiling product), satellite variables only, and the fused set of variables. It is clear from the results, the data mining method outperforms COAMPS<sup>TM</sup>-derived parameter in average error, and the benefit of using fused COAMPS<sup>TM</sup> and satellite variables is demonstrated.

#### **4.6 Conclusion and Future Work**

The initial results demonstrate the viability of using KDD to discover algorithms which can locate cloud ceiling presence, and calculate cloud ceiling height more successfully than numerical weather prediction models. Furthermore, results indicate that both COAMPS<sup>TM</sup> and satellite variables make contributions to the final results, indicating the value of a “fused data” approach. Still remaining is a study of night-time hours, and the other focus areas. Also, we are still in the process of applying these methods to rain rate and rain accumulation.

#### **Acknowledgment**

The support of the sponsor, the Office of Naval Research, under Program Element 0602435N is gratefully acknowledged. The assistance of the other members of the satellite team, past and present, at NRL Monterey (Code 7541), Bob Fett, Ron Englebretson, and Robin Brody of Scientific Applications International Corporation (SAIC), Marla Helveston of Anteon, Inc., and Melanie Wetzel of the Desert Research Institute is very much appreciated. The authors also wish to thank John Cook, Sue Chen, Jim Cummings, Teddy Holt, and other COAMPS researchers (all with NRL Monterey).

#### **References**

- [1] Duda, R.O., and P.E. Hart, 1973: *Pattern classification and scene analysis*. Wiley, 482 pp.
- [2] Hawkins, J.D, D.A. May, J. Sandidge, R. Holyer, and M.J. Heleveston, 1998: SSM/I-based tropical cyclone structural observations. Preprints, *9<sup>th</sup> Conference on Satellite Meteorology and Oceanography*, 230-233.
- [3] Welch, R.M, S.K. Sengupta, A.K. Goroch, P. Rabindra, N. Rangaraj, and M.S. Navar, 1992: Polar cloud and surface classification using AVHRR imagery: An intercomparison of methods. *J. Appl. Meteor.*, **31**, 405-420.
- [4] Key, J., 1990: Cloud cover analysis with arctic AVHRR data. Part II: Classification with spectral and textural measures. *J. Geophys. Res.*, **95**, 7661-7675.
- [5] Ebert, E., 1989: Analysis of polar clouds from satellite imagery using pattern recognition and a statistical cloud analysis scheme. *J. Appl. Meteor.*, **28**, 382-399.

- [6] Knottenberg, H., and E. Raschke, 1982: On the discrimination of water and ice clouds in multispectral AVHRR data. *Ann. Meteor.*, **18**, 145-14.
- [7] Allen, R.C., Jr., P.A. Durkee, and C.H. Wash, 1990: Snow/cloud discrimination with multispectral satellite measurements. *J. Appl. Meteor.*, **29**, 994-1004.
- [8] Tsonis, A.A., 1984: On the separability of various classes from the GOES visible and infrared data. *J. Climate Appl. Meteor.*, **23**, 1393-1410.
- [9] Garand, L., 1988: Automated recognition of oceanic cloud patterns. Part I: Methodology and application to cloud climatology. *J. Climate*, **1**, 20-39.
- [10] Baum, B.A., V. Tovinkere, J. Titlow, and R.M. Welch, 1997: Automated cloud classification of global AVHRR data using a fuzzy logic approach. *J. Appl. Meteor.*, **36**, 1519-1540.
- [11] Tag, P.M., R.L. Bankert, and L.R. Brody, 2000: An AVHRR multiple cloud-type classification package. *J. Appl. Meteor.*, **39**, 125-134.
- [12] Lee, J., R.C. Weger, S.K. Sengupta, and R.M. Welch, 1990: A neural network approach to cloud classification. *IEEE Trans. Geosci. Remote Sens.*, **28**, 846-855.
- [13] Bankert, R.L., 1994: Cloud classification of AVHRR imagery in maritime regions using a probabilistic neural network. *J. Appl. Meteor.*, **33**, 909-918.
- [14] Bankert, R.L., and D.W. Aha, 1996: Improvement to a neural network cloud classifier. *J. Appl. Meteor.*, **35**, 2036-2039.
- [15] Aha, D.W., and R.L. Bankert, 1995: A comparative evaluation of sequential feature selection algorithms. *Learning from Data: Artificial Intelligence and Statistics V*, D. Fisher and H.-J. Lenz, Eds., Springer-Verlag, 199-206.
- [16] Dvorak, V.F., 1984: Tropical cyclone intensity analysis using satellite data. NOAA Tech. Rep. NESDIS 11, U.S Dept. of Commerce, 47 pp.
- [17] Dvorak, V.F., 1975: Tropical cyclone intensity analysis and forecasting from satellite imagery. *Mon. Wea. Rev.*, **103**, 909-918.
- [18] Velden, C.S., T.L. Olander, and R.M. Zehr, 1998: Development of an objective scheme to estimate tropical cyclone intensity from digital geostationary satellite infrared imagery. *Weather and Forecasting*, **13**, 172-186.
- [19] Hawkins, J.D., T.F. Lee, J. Turk, C. Sampson, J. Kent, and K. Richardson, 2001: Real-time internet distribution of satellite products for tropical cyclone reconnaissance. *Bull. Amer. Meteor. Soc.*, **82**, 567-578.
- [20] May, D.A., J. Sandidge, R. Holyer, and J.D. Hawkins, 1997: SSM/I derived tropical cyclone intensities. Preprints, *22<sup>nd</sup> Conf. on Hurricanes and Tropical Meteorology*, 27-28.

- [21] Poe, G., 1990: Optimum interpolation of imaging microwave radiometer data. *IEEE Trans. Geosci. Remote Sens.*, **GE-28**, 800-810.
- [22] Ferraro, R.R., 1997: Special Sensor Microwave Imager derived global rainfall estimates for climatological applications. *J. Geophys. Res.*, **102**, 16715-16735.
- [23] Bankert, R.L., and P.M. Tag, 2002: An automated method to estimate tropical cyclone intensity using SSM/I imagery. *J. Appl. Meteor.*, **41**, 461-472.
- [24] Vislocky, R., and J.M. Fritsch, 1997: An automated observations-based system for short-term prediction of ceiling and visibility. *Weather and Forecasting*, **12**, 31-43.
- [25] National Weather Service, 1995: NGM-based MOS ceiling height guidance for the contiguous United States. *NWS Tech. Proc. Bull.*, **414**, 14 pp.
- [26] National Weather Service, 1990: NGM-based MOS ceiling height guidance for maximum/minimum temperature, probability of precipitation, cloud amount, and surface wind. *NWS Tech. Proc. Bull.*, **387**, 15 pp.
- [27] National Weather Service, 1988: LFM-based MOS ceiling amount guidance for the contiguous United States. *NWS Tech. Proc. Bull.*, **378**, 7 pp.
- [28] National Weather Service, 1981: The use of model output statistics for predicting ceiling, visibility, cloud amount, and obstructions to vision. *NWS Tech. Proc. Bull.*, **303**, 12 pp.
- [29] Weiss, S.M., and N. Indurkha, 1998: *Predictive Data Mining, a Practical Guide*. Morgan Kaufmann Pubs, Inc., San Francisco.
- [30] Berry, M., and G. Linoff, 1997: *Data Mining Techniques: For Marketing, Sales, and Customer Support*. John Wiley & Sons, Inc., New York, 454 pp.
- [31] Fayyad, U., D. Haussler, and P. Stolorz, 1996: Mining scientific data. *Communications of the ACM*, **39**, 51-57.
- [32] Fayyad, U., G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors, 1996: *Advances in Knowledge Discovery and Data Mining*. AAAI Press/MIT Press, Menlo Park, CA, 611 pp.
- [33] Weir, N., U. Fayyad, and S. Djorgovski, 1995: Automated star/galaxy classification for digitized POSS-II. *Astron. Journ.*, **6**, 2401-2412.
- [34] Burl, M.C., U. Fayyad, P. Perona, P. Smyth, and M.P. Burl, 1994: Automating the hunt for volcanoes on Venus. *Proceedings of Computer Vision and Pattern Recognition Conference (CVPR-94)*, IEEE Computer Science Press, 302-308.
- [35] Mesrobian, E., R. Muntz, E. Shek, S. Nittel, J. LaRouche, M. Krieger, C. Mechoso, J. Farrara, P. Stolorz, and H. Nakamura, 1996: Mining geophysical data for knowledge. *IEEE Expert*, **11**, 34-44.

- [36] Hadjimichael, M., P.M. Tag, R.L. Bankert, 1998: Discovering Model Bias in the Determination of Cloud Base Height. Proceedings, First American Meteorological Society Conference on Artificial Intelligence, Phoenix, AZ. January 11-16, J5-J8.
- [37] Hodur, R.M., 1997: The Naval Research Laboratory's coupled ocean/atmosphere mesoscale prediction system (COAMPS). *Mon. Wea. Rev.*, **125**, 1414-1430.
- [38] Wetzel, M.A., and L.L. Stowe, 1999: Satellite-observed patterns in stratus microphysics, aerosol optical thickness, and shortwave radiative forcing. *J. Geophys. Res.*, **104**, 31287-31299.
- [39] Lee, T.F., F.J. Turk, and K. Richardson, 1997: Stratus and fog products using GOES-8-9 3.9 micron data. *Weather and Forecasting*, **12**, 664-677.
- [40] Turk, J., C.S. Liou, S. Qiu, R. Scofield, M.B. Ba, and A. Gruber, 2001: Capabilities and characteristics of rainfall estimates from geostationary- and geostationary plus microwave-based satellite techniques. *Proceedings of Symposium on Precipitation Extremes: Prediction, Impacts, and Responses*. Amer. Meteor. Soc., 191-194.
- [41] Quinlan, J.R., 1993: *C4.5: Programs for Machine Learning*. Morgan Kaufmann Pubs, Inc., San Francisco, 302 pp.

# **Microwave Remote Sensing**

This page is intentionally left blank

## CHAPTER 11

### RECONSTRUCTION AND RESOLUTION ENHANCEMENT TECHNIQUES FOR MICROWAVE SENSORS

David G. Long

*Electrical and Computer Engineering Department  
459 Clyde Building, Provo, UT 84602  
E-mail: longee.byu.edu*

Microwave remote sensing instruments such as radiometers and scatterometers have proven themselves effective in a variety of Earth Science studies. The resolution of these sensors, while adequate for many applications, is a limiting factor to their application in other studies. As a result, there is a strong interest in developing ground processing methods which can enhance the spatial resolution of the data. A number of resolution enhancement algorithms have been developed based on inverse filtering and irregular sampling reconstruction. This Chapter discusses the use of resolution enhancement and reconstruction algorithms in microwave remote sensing. While the focus is on microwave instruments, the techniques and algorithms considered are applicable to a variety of sensors, including those not originally designed for imaging.

#### 1. Introduction

There are many types of remote sensing instruments, including optical, infrared, and microwave sensors. Microwave remote sensing instruments can be divided into two broad classes: passive (radiometers) and active (radars)<sup>55</sup>. Active microwave sensors can be further divided into four general classes: synthetic aperture radar (SAR) systems, scatterometers, altimeters, and weather radars. SAR systems are generally high resolution (100 m and finer) while spaceborne radiometers and scatterometers tend to be low resolution sensors (12 km to 75 km). The resolution of these latter sensors is suitable for the oceanic and atmospheric applications for which they were designed, but there is growing interest in applying such microwave sensor data to new applications requiring better resolution.

Further, while the next generation of spaceborne microwave sensors may have somewhat higher resolution, the extensive datasets of radiometer and scatterometer data offer an important baseline for studies of global change. This has resulted in interest in enhancing the resolution of historic microwave sensor data to facilitate comparison with higher resolution sensor data. To meet this need a number of algorithms for spatial resolution enhancement have been successfully developed and enhanced resolution microwave data is now being used operationally.

This Chapter considers the theory and methods of spatial resolution enhancement of microwave remote sensing data. While the focus is on spaceborne radiometers and scatterometers, the general reconstruction and enhancement theory discussed can be applied to a variety of sensors and applications, including creating images from sensors not originally designed for imaging. Section 2 provides background on the sensors considered. Section 3 provides background in resolution enhancement, contrasting inverse filtering, extrapolation, and reconstruction methods. The theory of irregular reconstruction is developed and algorithms for enhanced resolution reconstruction are considered. Section 4 considers the application of the technique to data from Earth Resources Satellite (ERS) -1 and -2 Active Microwave Instrument (AMI) scatterometer mode (hereafter termed ESCAT). Simulations are used to evaluate the effectiveness of the algorithms, along with actual data. A conclusion is provided in Section 5.

## 2. Spaceborne Microwave Sensors

A variety of active and passive microwave remote sensing instruments have flown in space. Some of these have collected long time series such as the Special Sensor Microwave Imager (SSM/I) radiometer<sup>18</sup>, flown on Defense Meteorological Satellite Program (DMSP) spacecraft since the early 1980's, ESCAT<sup>2</sup> operating from 1982, and the SeaWinds instrument<sup>50</sup> operating aboard QuikSCAT since 1999. Prior sensors include the NASA Scatterometer (NSCAT)<sup>38</sup> which operated in 1996 and 1997, the Seasat Scatterometer (SASS)<sup>23</sup> and Multichannel Microwave Radiometer<sup>39</sup> both in 1978, and the Nimbus radiometer series operating in the late 1970's, among others. Together these instruments have demonstrated the utility of microwave sensors in the study and monitoring of the Earth's land, ocean, and atmosphere. The global coverage, but low resolution, of these sensors complements the high resolution, but limited coverage, of SAR systems.

Radiometers are passive, receive-only sensors which measure the thermal emission (brightness temperature) of the target in the microwave band<sup>55</sup>. The apparent scene brightness temperature is related to the emissivity and temperature of the surface and is modified by the intervening atmosphere. By appropriate selection of operating frequencies in several microwave bands, the temperature and moisture content of the atmosphere<sup>22</sup>, as well as key surface properties such as land surface temperature<sup>37</sup>, soil and plant moisture<sup>20,42</sup>, sea-ice mapping<sup>54</sup>, snow cover classification<sup>16</sup>, and wind speed (over the ocean)<sup>57</sup>, can be retrieved. Radiometer data is being operationally used in weather forecasting and sea-ice monitoring.

Scatterometers are real aperture radars that operate by transmitting a pulse of microwave energy towards the Earth's surface and measuring the reflected energy. The backscattered energy is related to the normalized radar cross-section ( $\sigma^{\circ}$ ) via the radar equation<sup>55</sup>. The spatial response function of the sensor determines the spatial resolution of the  $\sigma^{\circ}$  observation, with typical resolutions varying from 25 to 50 km. Originally designed for retrieval of near-surface winds over the ocean,



scatterometer data is also being applied to the study of tropical vegetation, polar ice, and global change<sup>26,29,33,34,59</sup>. Like radiometer data, scatterometer data is operationally used in weather forecasting and sea-ice monitoring.

### 2.1. Radiometers

A radiometer measurement is the integral of the product of the scene brightness and the antenna pattern. The  $i^{\text{th}}$  measurement  $T_a(i)$  (in  $K$ ) is obtained by integrating the product of surface brightness response  $T_b(x, y)$  and the antenna gain pattern at the surface  $G_i(x, y)$ ,<sup>13,55</sup>

$$T_a(i) = \bar{G}_i^{-1} \iint G_i(x, y) T_b(x, y) dx dy, \quad (1)$$

where

$$\bar{G}_i = \iint G_i(x, y) dx dy. \quad (2)$$

The integrals are over the surface area corresponding to the non-negligible gain of the antenna. The dependence of  $G$  on  $i$  arises from the boresight pointing of the antenna which changes as the antenna scans the surface. Note that the antenna pattern acts as a low pass filter of the surface brightness, limiting the effective spatial resolution of the measurement to approximately the 3 dB beamwidth. Radiometer measurements are “noisy” due to the limited integration time available for each measurement.

### 2.2. Scatterometers

A radar scatterometer is designed to determine the normalized radar cross section ( $\sigma^\circ$ ) of the surface. The primary application of spaceborne scatterometers have been the measurement of near-surface winds over the ocean. By combining  $\sigma^\circ$  measurements from different azimuth angles, the near-surface wind vector over the ocean’s surface can be determined using a geophysical model function<sup>23,38</sup> which relates wind and  $\sigma^\circ$ . The scatterometer directly measures  $\sigma^\circ$  via measuring the backscattered power from a transmitted pulse. Due to thermal noise in the receiver, radiometric noise and speckle, the power measurement is corrupted by noise. A separate measurement of the noise-only power is subtracted from the signal+noise measurement to yield the backscattered power “signal” measurement  $P_S$ . The observed  $\sigma^\circ$  is then computed using the radar equation<sup>55</sup>. Ignoring the incidence angle dependence of  $\sigma^\circ$ , the radar equation can be approximately expressed as

$$P_S = \iint \frac{P_T G(x, y) \lambda \sigma^\circ(x, y)}{(4\pi)^3 R^4(x, y)} dA = \iint h(x, y) \sigma^\circ(x, y) dA \quad (3)$$

where  $P_T$  is the transmit power,  $G(x, y)$  is the antenna gain pattern on the surface,  $dA$  is the differential area,  $R(x, y)$  is the slant range,  $\sigma^\circ(x, y)$  is the surface  $\sigma^\circ$ ,  $h(x, y)$  is the equivalent spatial response function, and the integrals are over the

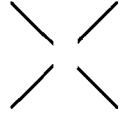
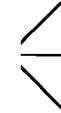
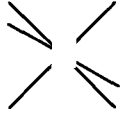
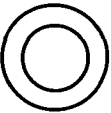
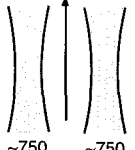
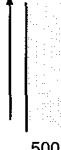
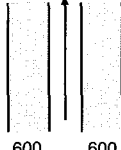
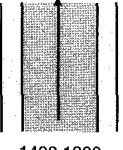
	SASS	ESCAT	NSCAT	SeaWinds
FREQUENCY	14.6 GHz	5.3 GHz	13.995 GHz	13.6 GHz
ANTENNA AZIMUTHS				
POLARIZATIONS	V-H, V-H	V ONLY	V, V-H, V	V-OUTER/H-INNER
BEAM RESOLUTION	FIXED DOPPLER	RANGE GATE	VARIABLE DOPPLER	PENCIL-BEAM
SCIENCE MODES	MANY	SAR, WIND	WIND ONLY	WIND/HI-RES
RESOLUTION ( $\sigma^\circ$ )	50 km	50 km	25 km	Egg: 25x35 km Slice: 6x25km
SWATH, km				
INCIDENCE ANGLES	0° - 70°	18° - 59°	17° - 60°	45° & 54.4°
DAILY COVERAGE	VARIABLE	< 41 %	78 %	92 %
MISSION DATES	SEASAT: 6/78 — 10/78	ERS-1: 92—96 ERS-2: 95—01(?)	ADEOSI: 8/96 — 6/97	QuikSCAT: 6/99 — ADEOSII: ~11/02

Fig. 1. Scatterometer Comparison

surface area corresponding to the non-negligible gain of the antenna and/or signal processing filters. Given  $P_S$ , the observed  $\sigma^\circ$  is computed as  $\sigma^\circ = X^{-1}P_S$  where

$$X = \iint \frac{P_T G(x, y) \lambda}{(4\pi)^3 R^4(x, y)} dA = \iint h(x, y) dA. \quad (4)$$

A summary comparison of the wind scatterometers that have flown in space is shown in Fig. 1. SASS, NSCAT and ESCAT used a fan-beam antenna configuration while SeaWinds employs a dual rotating pencil-beam antenna. In a fan-beam scatterometer, along-track resolution is obtained by a combination of a narrow antenna pattern and the timing of transmit pulses integrated into a single measurement cell. Cross-track resolution is obtained either by range gate filtering (ESCAT) or by Doppler filtering (SASS and NSCAT)<sup>38</sup>. ESCAT is described in greater detail later.

While a scanning scatterometer collects measurements at a constant incidence angle<sup>50</sup>, fan-beam scatterometers measure  $\sigma^\circ$  at a variety of incidence angles. Since the target response varies with incidence angles, a model for the incidence angle dependence of the target response is used to generate normalized images. Over most natural surfaces within the incidence angle range  $20^\circ \leq \theta \leq 60^\circ$ , corresponding to the range of scatterometer measurements, a linear model for  $\sigma^\circ$  (in dB) as a function

of incidence angle can be used<sup>a</sup>, i.e.  $\sigma_{\text{dB}}^{\circ}(\theta) = \mathcal{A} + \mathcal{B}(\theta - 40^{\circ})$  where  $\mathcal{A}$  is the  $40^{\circ}$  incidence angle-normalized  $\sigma^{\circ}$  and  $\mathcal{B}$  is the dependence of  $\sigma^{\circ}$  on the incidence angle  $\theta$ . The  $\mathcal{A}$  and  $\mathcal{B}$  coefficients are functions of the geophysical properties of the surface. Note that  $40^{\circ}$  is the approximate center incidence angle over the swath and is a convenient angle for making comparative analyses.

As previously noted, scatterometer measurements are noisy. The measurement accuracy is frequently expressed in terms of the normalized standard deviation, or  $K_p$  of the measurements<sup>36</sup>.  $K_p$  is sometimes known as the “scatterometer radiometric accuracy”. For ERS-1  $K_p(k)$  is approximately 5%. For SASS, NSCAT, and SeaWinds  $K_p(k)$  varies from as low as 1% to 15%, though it is sometimes higher.

Reconstruction and resolution enhancement methods can be applied to other real aperture radars as well. For example, the Tropical Rain Mapping Mission (TRMM) Precipitation Radar<sup>24</sup>, while it employs range resolution to map the vertical profile of rain, it also makes real-aperture surface backscattering measurements which can be applied in scientific studies<sup>8,30,51</sup>.

### 3. Reconstruction and Resolution Enhancement

The resolution of scatterometers and radiometers is adequate for ocean applications but is too coarse for many land and ice applications. However, because of frequent global coverage, they are desirable candidates for resolution enhancement algorithms. Since the data from these instruments is used in geophysical studies, accuracy is crucial in resolution enhancement. Further, an improvement in the actual *effective* resolution of the data is expected from such algorithms. While it is tempting to interpolate the available data onto a high resolution grid in an attempt to make the pixel size (sometimes called the *pixel resolution*) finer, this does not improve the *effective* resolution of the resulting image. While various definitions of effective resolution exist<sup>21,25</sup>, a common working definition is the resolving capability for two closely spaced objects. The objects are considered individually “resolved” if there is a 3 dB change in image value between them against a high contrast background. The gap between the objects defines the effective resolution.

Algorithms for spatial resolution enhancement can be divided into three broad categories: extrapolation, ad hoc techniques, and reconstruction. Extrapolation algorithms can be further divided into two classes: pure extrapolation and multi-channel extrapolation. The former includes algorithms which use maximum entropy to extrapolate techniques the signal spectrum. Based on our common experience in polynomial extrapolation, extrapolation must be used with caution since it can

---

<sup>a</sup>While not applicable for all targets, over the Amazon Rainforest (which exhibits high volume scattering) a “gamma” normalization may be used with some success, i.e.,

$$\gamma(k) = \sigma^{\circ}(k) / \cos \theta(k)$$

where  $\theta(k)$  is the incidence angle of the  $k^{\text{th}}$  measurement of  $\sigma^{\circ}$ . Over the incidence angle range  $[20^{\circ}, 60^{\circ}]$ ,  $\gamma(k)$  is approximately constant.

produce misleading results.

Multi-channel extrapolation algorithms have been used with success with microwave data<sup>3,48</sup>. These algorithms rely on the collection of observations at several different frequency channels which have differing spatial resolutions. The goal of the algorithm is to extrapolate the signal characteristics of the coarse resolution channels to be commensurate with the resolution of the fine resolution channels based on the correlation in the target response between the channels. As can be expected, decorrelation from target variability and modeling uncertainty between the channels is a key limitation of this approach.

A number of image restoration and ad hoc enhancement techniques have been developed<sup>14,47,48,49,58</sup>. An example of an ad hoc technique is simple linear interpolation to increase the pixel density followed by noise addition. The perceptual “resolution” of the interpolated image is improved by adding white noise to the image. This results in the image appearing to have more higher spatial frequency information than is actually supported by the underlying data.

Reconstruction algorithms rely on reconstructing the original signal based on sampled observations. A classic reconstruction algorithm is the well-known Nyquist uniform sampling theorem. Given uniformly-spaced ideal samples of a band-limited signal, the original signal can be exactly reconstructed by ideal low pass filtering of the samples so long as the sample spacing is at least twice the highest frequency present in the signal. Reconstruction algorithms become more complicated with the introduction of irregular sampling and variable apertures, a common problem in microwave remote sensing<sup>9</sup>.

We note that in the signal and image processing literature, the term “image enhancement” generally refers to inverse filtering techniques: to enhance a pre-existing image, an inverse filter is applied by convolution. This approach generally requires a constant (over the image to be processed) aperture function and a pre-existing, uniformly sampled image. The general theory for this approach is well-known<sup>21,25</sup> and the literature is replete with examples of variations of this method, including techniques to estimate the aperture function from the image. The approach has been successfully applied in remote sensing<sup>3,4</sup>. As noted, many microwave sensors do not produce data on a uniform grid, and the data must be converted to an image prior to applying such methods. Simple gridded images can be generated with the widely used “drop-in-the-bucket” technique by assigning each measurement to a grid element in which its center falls, or some variation thereof. However, the effective resolution of such images is dictated by the aperture response rather than the grid spacing.

In effect, using reconstruction techniques creates optimal images and performs the function of “image enhancement” at the same time; hence our interest in reconstruction-based techniques. As a result, the remainder of this Chapter focuses exclusively on reconstruction based techniques. Irregular sampling and reconstruction theory are emphasized to support the often non-rectilinear grid sampling of

microwave sensors.

### 3.1. Microwave Sensor Sampling and Resolution Enhancement

Typical microwave sensor observations can be modeled as an array of samples of spatially filtered surface data. The aperture function for each measurement is defined by the sensor antenna pattern and/or the signal processing techniques used to resolve the antenna illumination pattern into smaller spatial elements, e.g. Doppler filtering or range gating. Spatial sampling is typically obtained via pulsed operation and antenna scanning. While ideally such sampling is on a regular sampling grid, this is not the case for past and present sensors which generally have irregular or varying sampling grids and spatially varying aperture functions. Some sensors (e.g., SASS) can not even be considered ‘imaging sensors’ since the aperture filtered samples do not completely cover the surface for a single pass.

Resolution enhancement algorithms provide improved resolution images by taking advantage of oversampling and the response characteristics of the aperture function to reconstruct the underlying surface function sampled by the sensor. The goal of the algorithm is to generate images from the observations at an effective spatial resolution better than the 3 dB resolution of the sensor; hence the term “resolution enhancement”. When single-pass sampling has inadequate sampling density, multiple observation passes can be combined to improve the sampling density, producing the required oversampled observations for spatial resolution enhancement at the cost of reduced temporal resolution<sup>3,9,35</sup>.

### 3.2. Irregular Sampling and Reconstruction

Let  $f(x, y)$  represent the true surface image (e.g.,  $\sigma^\circ$  or brightness temperature) at a location  $(x, y)$ . The measurement system is modeled by

$$z = Hf + \text{noise} \quad (5)$$

where  $H$  models the measurement system (including the sample spacing and the system’s spatial response function, hereafter termed the aperture response function) and  $z$  is the vector of observations made by the sensor. The measurements  $z$  are a discrete sampling of  $f$  convolved with the aperture function (which may be different for each measurement). An individual measurement  $z_i$  can be written as

$$z_i = \iint h_i(x, y) f(x, y) dx dy + \text{noise} \quad (6)$$

where  $h_i(x, y)$  is the aperture response function of the  $i^{\text{th}}$  measurement. The aperture response is also called the point-spread function.  $h_i(x, y)$  is a function of the antenna pattern and the effective signal processing filter response for the  $i^{\text{th}}$  measurement, see Eqs. (1) and (3).

Reconstruction and resolution enhancement involves inverting Eq. (5)

$$\hat{f} = \hat{H}^{-1} z \quad (7)$$

where  $\hat{f}$  is the estimate of  $f$  derived from the measurements  $z$ . The inverse of  $H$ ,  $\hat{H}^{-1}$ , is exact if  $H$  is invertible. If  $H$  is not invertible, an approximate solution must be used. The low-pass nature of typical aperture functions passes low spatial frequencies, but attenuates and may even null out higher spatial frequencies. Full reconstruction from sufficiently dense sampling can be considered resolution enhancement since high frequency information suppressed (but not nulled out) by the aperture function is recovered.

Because the sensor measurements are noisy, a tradeoff between resolution enhancement and the noise level in the reconstructed signal exists since high frequency noise tends to be amplified along with the signal in the reconstruction process.

### 3.3. Sampling and Reconstruction

The traditional approach to sampling and reconstruction is founded on uniform sampling and the well-known Nyquist sampling theorem: a low pass (band limited) function can be completely reconstructed from regularly spaced samples if the sample rate exceeds twice the maximum frequency present in the signal<sup>41</sup>. In typical application, signal reconstruction from the samples is accomplished with only a low pass filter and the aperture function is treated as an ideal low pass filter ignored in the reconstruction. For this case, the recovered frequencies are deemed limited to 1/2 the sampling frequency or the cutoff frequency (e.g. the 3 dB rolloff point) of the aperture function, depending on which is lower. The aperture function filters out high frequency components of the signal that might otherwise cause aliasing in the reconstructed signal.

Since the aperture function of a microwave sensor is the result of the antenna pattern and signal processing, it has side lobes. The resulting measurements thus contain information regarding higher frequency components of the original signal. If the (possibly irregular) sampling is sufficiently dense, this information can be recovered by inverting the effects of both the aperture function and the sampling. The reconstruction compensates for the aperture filtering by amplifying attenuated frequencies, though the aperture function may limit the reconstruction due to nulls in its spectrum.

If the sampling is regular (uniform) with a fixed aperture function, reconstruction can be accomplished with low pass filtering and Wiener filtering, a well-known inverse filtering technique that also accounts for noise in the measurements<sup>41</sup> (see, for example, Alvarez-Perez *et al.*<sup>1</sup> for an application of such a technique to ERS scatterometer data). However, inverse filter methods are difficult to apply when the sample spacing is irregular or when the aperture functions vary between different observations. Instead, irregular reconstruction methods must be applied.

While the theory of uniform sampling and reconstruction is well-known, irregular sampling and reconstruction theory is much less familiar. Here we review the general theory for irregular sample reconstruction.

As in uniform sampling, the sample spacing, or *sampling density*, limits the

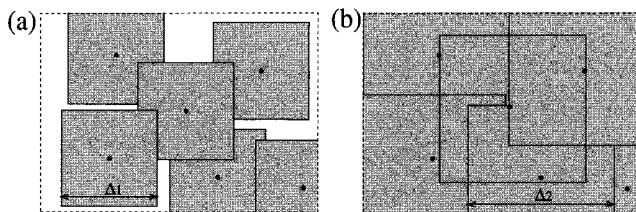


Fig. 2. Graphical illustration of  $\delta$ -dense in 2-D. A box of size  $\Delta$  is centered at each sample point. (a) With  $\delta = (\Delta_1, \Delta_1)$ , the union of the boxes around each sample point is too small to cover the image space. (b)  $\delta = (\Delta_2, \Delta_2)$  is sufficiently large.  $\delta$ -dense corresponds to the smallest  $\delta$  which covers the full image [from Early and Long (2000)].

signal reconstruction. A simple method for parameterizing the sampling density for an arbitrary irregular grid is based on  $\delta$ , the maximum sample spacing. A formal definition and discussion of  $\delta$ -dense sampling is provided in Early and Long<sup>9</sup>; here a less formal approach is adopted. In one dimension,  $\delta$  specifies the maximal spacing of the samples. In two dimensions  $\delta$ -dense is defined as the minimum sized rectangle centered at each sample point such that the union of the boxes completely fills the image space (see Fig. 2).

Gröchenig<sup>15</sup> derives a relationship between the sample grid parameter  $\delta$  and the recoverable frequencies (a band limited frequency range denoted by  $\Omega = [\omega_1, \omega_2]$ ) of the original signal, showing that the signal can be completely reconstructed if

$$\delta \cdot \omega = \sum_{i=1}^2 \delta_i \omega_i < \ln(2). \quad (8)$$

If the spectrum of the original signal has a region of support  $\Omega = [-\omega_0, \omega_0]^2$ , and the  $\delta$ -dense sampling grid has  $\delta_1 = \delta_2$ , the sampling density must satisfy

$$\delta_1 < \frac{\ln(2)}{2\omega_0}. \quad (9)$$

This requires that the minimum irregular sampling density must be higher than the Nyquist uniform sampling density, i.e.,  $1/\ln(2) \approx 1.44$  times the Nyquist rate for uniformly spaced samples. This ‘oversampling’ is required to ensure reconstruction from the irregular sampling grid.

Thus, for irregular sampling, Gröchenig’s theory is equivalent to the well-known Nyquist theory of sampling and reconstruction for uniform sampling: for complete reconstruction (1) the original signal must be bandlimited or aliasing and information loss results and (2) the sampling must be sufficiently dense. While the original signal can be recovered only if the maximum frequency is less than sampling density, the aperture function used to create the samples can introduce information loss, restricting the frequencies which can be reconstructed.

### 3.4. Reconstruction Algorithms

Gröchenig's proof is constructive, with an algorithm for reconstructing the original signal from the samples (in effect, inverting  $H$ ). As demonstrated by Early and Long<sup>9</sup> Gröchenig's algorithm is equivalent to block additive Algebraic Reconstruction Technique (ART) when used with a suitably defined operator  $H$  based on the sampling, the aperture function and the signal bandwidth. ART methods have been extensively studied<sup>6,12,17</sup> and a number of practical numerical algorithms have been developed. Such methods thus form a basis for the practical reconstruction of irregularly sampled signals in remote sensing.

As previously noted, the sensor observations or measurements can be viewed as ideal samples of an aperture filtered image where the aperture filtered image is the true image convolved with an aperture function. In general, each observation can use a different aperture function. For a given aperture function, nulls in the frequency response of the aperture function result in lost information. For a single aperture, this information is permanently lost and cannot be recovered via reconstruction. However, when multiple aperture functions are used, a net effective aperture function can be defined from the appropriately averaged individual measurement aperture functions<sup>9</sup>. Nulls in the effective aperture function correspond to the intersection of the nulls of individual aperture functions. So long as the sampling density requirements are met for the remaining frequencies, only the frequencies corresponding to the nulls in the net effective aperture function are lost. All other frequencies can be recovered in the reconstruction, subject to the sampling considerations. Though information in spectral nulls of the aperture function is permanently lost, some ART-based reconstruction algorithms can "fill-in" data for missing frequencies based on particular mathematical criteria.

Here we develop the ART algorithm. For convenience the original signal is treated as discrete with uniformly-sized pixels, but at a very fine scale (much smaller than the sample spacing). Each measurement or observation  $s_i$  covers a number of these small pixels [compare Eq. (6)]

$$s_i = \sum_j h_{ij} a^j \quad (10)$$

where  $a^j$  are elements of the vector  $a$  of row-scanned image pixels of the true signal image and  $h_{ij}$  is the effective aperture response function for the  $i$ th measurement on the  $j$ th pixel. The sum is computed over the pixels for which  $h_{ij}$  is non-negligible.

Block additive ART (AART) can be written as<sup>17</sup>

$$a_{n+1}^j = a_n^j + \frac{\sum_i (s_i - p_i) h_{ij}}{\sum_i h_{ij}} \quad (11)$$

where  $a_n$  is the  $n^{\text{th}}$  iterative estimate of  $a$ , and  $p_i$  is the back projection

$$p_i = \sum_j h_{ij} a_n^j \quad (12)$$



corresponding to the  $i$ th measurement at the  $n^{\text{th}}$  iteration. In effect, all measurements that cover the pixel of interest are summed and normalized to create the per-pixel update.

The update on the right side of Eq. (11) is a function of the measurement vector  $s$  and the back projection vector  $p$  computed from the  $n^{\text{th}}$  iterative estimate. The vector  $s$  is the sampled convolution of the true image and the aperture function(s), expressed in matrix form as  $s = Ha$  where  $H$  (with elements  $h_{ij}$ ) is the sampled aperture function for each measurement. Then, Eq. (11) becomes (noting  $p = Ha_n$ )

$$\begin{aligned} a_{n+1} &= a_n + H'(s - p) \\ &= a_n + H'(Ha - Ha_n) \\ &= a_n + \mathcal{H}(a - a_n) \end{aligned} \tag{13}$$

where the  $a$ 's are row-scanned image vectors,  $H'$  is the row-normalized transpose of  $H$  with elements  $h_{ji}/\sum_k h_{kj}$ . To perform reconstruction consistent with the  $\delta$ -dense sampling, a low pass filter is applied to the rows of  $H$ . The resulting  $\mathcal{H}$  is invertible over the frequency range defined by the aperture function and sampling<sup>9</sup>.

As previously noted, noise in the measurements tends to be amplified along with the desired signal. In Wiener filtering, the reconstruction filter response is modified so that when a specified noise-to-signal ratio threshold is exceeded, the response is set to zero to minimize noise amplification<sup>25</sup>. A similar approach can also be used to modify the rows of  $H$  in the reconstruction.

Since the reconstruction algorithm is iterative, computational considerations may limit the number of iterations, resulting in a less-than-optimal reconstruction. Thus, there is a tradeoff between the reconstruction accuracy and resolution and the number of iterations. Increasing iterations improves the resolution. Truncation of the iterations can also be considered a method of regularization<sup>25</sup>.

In general, iterative reconstruction suffers from two forms of error: reconstruction error and noise amplification. The former is the difference between the iterative image estimate and the noiseless true image. Noise amplification results from the inverse filtering since the reconstruction algorithm acts as a high pass filter. Excess measurements (due to over sampling or repeated observations) contribute to an improvement in the signal to noise ratio of the estimated image due to averaging in the reconstruction algorithm. Thus, increasing the number of measurements improves the noise level, even if the effective sampling density is not increased.

To avoid having to solve for and explicitly compute within the space delineated by the aperture function, regularization techniques can be used to compute a unique solution on the full space. The AART algorithm includes least squares regularization, though a variety of regularization schemes can be applied to generate an estimate of the signal.

As noted by Early and Long<sup>9</sup> AART and multiplicative ART (MART) solutions differ only by regularization implicit in the algorithms. AART is equivalent to a least squares estimate in the limit based on the minimization of  $\|x^2\|$  subject to  $y = Hx$

while MART with damping maximizes the signal entropy  $-\sum_{j=1}^n x_j \ln x_j$  with the same constraint. The AART estimate is strictly contained within the band limited space defined by the aperture function and sampling, while the MART estimate is not confined to this space – additional frequency content in the null space may be added by the algorithm to create a sharper image<sup>21</sup>, subject to the constraint  $y = Hx$ . The difference between the AART and MART solutions (in the iteration limit) is contained in the null space of  $\mathcal{H}$ .

Alternate variations of these reconstruction algorithms can also be used effectively. The Scatterometer Image Reconstruction (SIR) algorithm is a derivative of MART developed for scatterometer image reconstruction<sup>35</sup>. It includes a non-linearity in the update to minimize the effects of noise on the reconstruction and is preferred over MART<sup>9,58</sup>. The SIR algorithm has been widely applied to both scatterometer and radiometer data<sup>29,31,35</sup>. When applied to scatterometer data, the multivariate form of SIR estimates the incidence angle dependence of the scatterometer data<sup>35</sup> and is used below. Another reconstruction approach is based on the Backus-Gilbert technique<sup>13,46,52,53</sup>, contrasted with SIR by Long *et al.*<sup>31</sup>.

While the overall performance of AART, MART and SIR algorithms are similar, at lower reconstruction errors MART and SIR have lower noise amplification than AART, and at the lowest reconstruction errors, SIR has the lowest noise. Thus, SIR is more robust in the presence of noise, particularly at low signal to noise ratios<sup>9</sup>. Further, the subjective image quality for SIR at a given reconstruction error level is better than corresponding MART or AART products when used with scatterometer data. The ultimate limits to resolution enhancement are the sampling density, nulls introduced by the aperture function(s), the acceptable noise level, and the temporal stability of the study area<sup>9,35</sup>. Inverse filtering of the reconstructed SIR image can further improve the quality of the image<sup>7</sup>.

#### 4. Application Example: ESCAT Resolution Enhancement

As has been noted, reconstruction-based resolution enhancement is based on restoring attenuated information in the sidelobes of the spatial response function within the support of the sampling. High side lobes in the spatial response make this easier, though information can be recovered even from sensors with low sidelobes. To illustrate this we consider a particular application example: ESCAT, which uses a processing window designed to minimize sidelobes. ESCAT also has a much narrower swath than other sensors and thus requires many more passes over the target to achieve a similar high sampling density. Application of resolution enhancement techniques to ESCAT is thus more demanding than for other scatterometers<sup>9,10,29,35</sup>.

For ESCAT, ground processing is used to spatially filter and resample the raw instrument measurements. Several pulses corresponding to each along-track cell are integrated into a single “50 km” resolution measurement. Nominally 50 km  $\sigma^\circ$  measurements are reported on a 25 km grid for each antenna. A spatial smoothing filter (Hamming window) is applied when integrating the pulses. This filter smoothes the

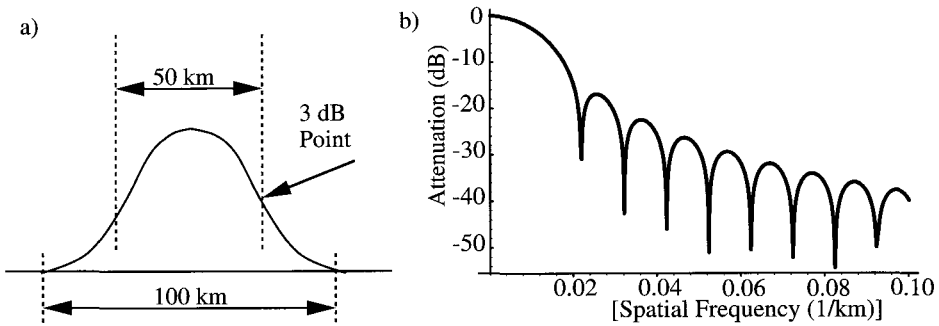


Fig. 3. a) Plot of the ESCAT response along a section of the spatial resampling filter. b) Spatial spectrum of a).

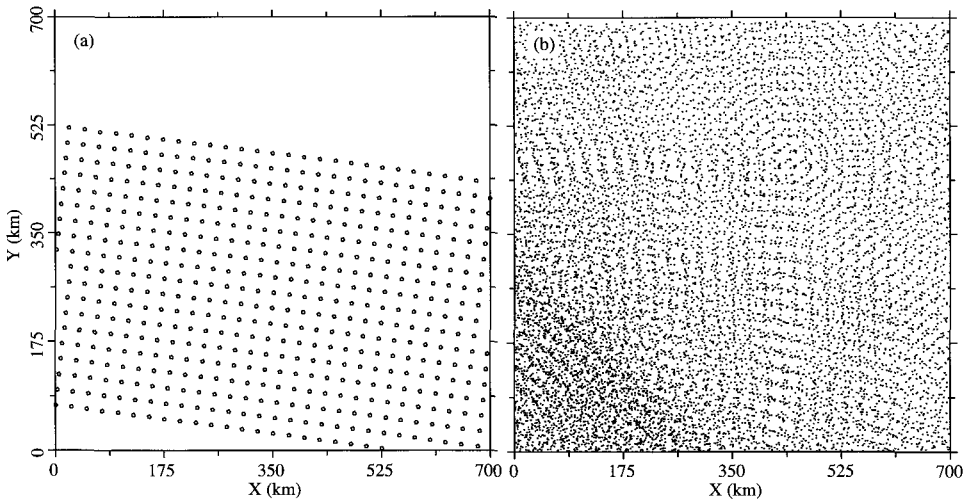


Fig. 4. a) Locations of the  $\sigma^0$  measurement centers for a single pass over a small study area. b) Locations of the  $\sigma^0$  measurement centers over a 6 day period.

$\sigma^0$  response so that it corresponds to the average  $\sigma^0$  response for a 50 km circle (see Fig. 3)<sup>b</sup>. This weighting function, chosen to minimize sidelobes and aliasing, is the primary factor in determining the effective measurement response or aperture function. The windowed spatial resampling has the desirable dual effects of 1) reducing the noise level (i.e., decreasing  $K_P$  and improving the radiometric resolution) and 2) compensating for the varying areas and resolutions of the individual measurements. For each 25 km grid element there are three measurements of  $\sigma^0$ , one from each antenna beam. In many land and ice applications these may be combined; however, care must be used for surfaces exhibiting azimuthal variation in  $\sigma^0$ .

<sup>b</sup>E. Attema, Personal communication.

Based on the Nyquist theorem, frequencies up to  $0.02 \text{ km}^{-1}$  ( $1/50$ ) can be reconstructed from uniform 25 km sampling. If multiple passes over a given study region are combined, the effective sampling density can be improved. Combining multiple passes results in a dense, but irregular sampling grid. For example, Fig. 4a illustrates the ESCAT measurement locations for a single pass over a particular study area. These are on a 25 km uniformly spaced sampling grid. Figure 4b shows the locations of all measurements collected over the study area in a 6 day period. Note that NSCAT and SeaWinds achieve significantly denser sampling during a similar period<sup>9</sup> (see also Fig. 1). Of course, in order to usefully combine the multiple passes to achieve the dense sampling the following assumptions must be made: 1) the instrument calibration is stable, 2) the surface  $\sigma^\circ$  remains essentially constant for the combination period, 3) the surface  $\sigma^\circ$  does not vary with azimuth angle since different passes may observe the surface at different azimuth angles, and 4) the location and response function(s) of the measurements are accurately known<sup>35</sup>.

To address these considerations we note that both ERS-1 and ERS-2 scatterometers have demonstrated excellent calibration stability<sup>32</sup>, satisfying assumption 1. Applying assumption 2 limits the multiple pass combination technique to stationary or very slowly evolving targets such as land<sup>10</sup>; ocean or rapidly moving sea ice regions are unsuited for combining multiple passes. Assumption 3 can be applied for much of the Earth's surface, although there are known regions of Antarctic glacial ice which exhibit significant azimuth dependence in C-band  $\sigma^\circ$  <sup>11,19,26,27,28,45</sup> and caution must be exercised in such cases. Given care in the implementation of the ground processing of the ESCAT data, assumption 4 is reasonable.

Proper reconstruction involves inverting the effects of the sampling and aperture function (i.e. the resampling window for ESCAT) over the frequency range supported by the sampling. Windowing introduces nulls in the measurement spectra (see Fig. 3) and at such frequencies the original signal cannot be recovered. However, over frequencies supported by the sampling at which the signal spectra is merely attenuated, the original signal can be completely recovered: the reconstruction algorithm can compensate for the attenuation introduced by the resampling filter—even in the highly attenuated sidelobe regions. Simulations demonstrate that sidelobe compensations of over 60 dB are possible using SIR with this aperture, sufficient sampling, and long enough iteration.

We note that the original SIR algorithm was developed for SASS where the aperture function could be approximated by a boxcar or rect function, simplifying the algorithm<sup>35</sup>. However, for the SSM/I and ESCAT it is appropriate to use the actual response function in the SIR algorithm. As a general rule, the lower sidelobes of the ESCAT aperture function require more iterations to achieve the same level of resolution enhancement compared to more rect-like aperture functions.

The noise performance of SIR can optionally be improved via use of a median filter, a modification known as SIRF (SIR with Filtering)<sup>35</sup>; however, this has the side effect of reducing the effective resolution. SIRF was used with SASS measure-

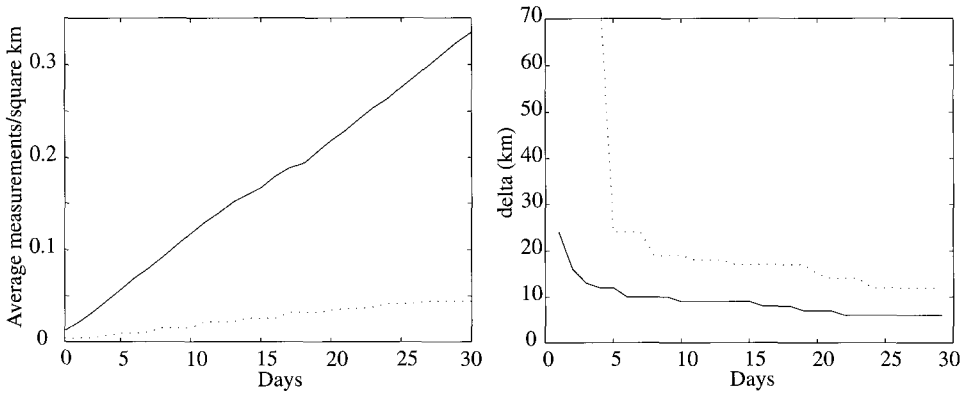


Fig. 5. a) Plot of  $\delta$ -dense versus time for two study areas. b) Average number of  $\sigma^\circ$  measurements per square km as a function of time for two locations. In both plots the solid line is for a polar location while the dotted line corresponds to an arbitrary mid-latitude location.

ments due to their high noise level. ESCAT measurements are much less noisy than SASS measurements due to the better SNR and so the median filter is not used with ESCAT. A previous author<sup>56</sup> erroneously stated that resolution enhancement can not be applied to ESCAT data due to the windowing applied in the measurement process. However, as we have shown the windowing primarily only degrades the signal to noise ratio and information in even very low sidelobes can be recovered if desired and the noise enhancement can be tolerated. The same author attributed the resolution enhancement of SIR to the use of a median filter; however, SIR does not include a median filter. The primary limitations of ESCAT resolution enhancement are the degradation of the signal to noise ratio and the computational time, which limits the number of iterations.

#### 4.1. Sampling Density

We now consider the sampling density achievable for ESCAT. Combining multiple passes increases the sampling density, quantified by  $\delta$ . However, due to the ERS orbit and swath geometry, the number of passes in a given time period and their relative orientation and spacing varies considerably over different regions of the Earth. Further, since the scatterometer mode can not be used when SAR data is being collected, there are gaps and missing data. In any case, the multi-orbit sampling is suboptimum. Nevertheless, as the number of overpasses is increased,  $\delta$  decreases in a location-dependent manner. To illustrate the relationship between the  $\delta$  parameter and time, Fig. 5 plots  $\delta$  computed over two study regions, one at mid latitudes in the Northern Hemisphere and the other in the polar region, which gets more frequent coverage. The value of  $\delta$  and the average number of measurements per square km are both shown. A tradeoff between  $\delta$  and time is apparent. There is a general linear trend in the number of measurements while  $\delta$  exhibits an

exponential decay with time. We note that the reconstruction resolution limit is approximately  $2.9 \times \delta$  (see Eq. 9). Because of concerns about temporal variations in the surface, we desire to minimize the temporal period of the measurements to be combined. However, to achieve a given sampling density (and therefore resolution in the resulting reconstructed image), a minimum time interval is required, termed the ‘imaging period.’

It should be clear that the selection of the imaging period depends on the intended application. However, the point of inflection in the  $\delta$  versus time curve provides a good tradeoff between length of the imaging period and sampling density. Examining Fig. 5 we see that six days provides a  $\delta$  of 10 km (corresponding to an effective reconstruction resolution of approximately 29 km) for the polar region while nearly a month is required for the particular mid latitude region evaluated. For the same period a  $\delta$  of approximately 6 km near the poles is achieved. Higher density sampling is possible but requires longer time periods.

We note that  $\delta$  is dependent on the *worst-case* sample spacing over the study area, which may occur at only one point; the sample density elsewhere is higher. While  $\delta$ -dense defines the sample spacing required to guarantee the reconstruction resolution everywhere, regions of the image with denser sampling can yield higher enhancement in practice due to locally denser sampling. For example, the area covered by a single pass has a  $\delta$  of 25 km, even if the remainder of the image area is not covered at all. (This explains the high starting value of  $\delta$  in Fig. 5.) Thus, the reconstructed resolution can vary over the image.

#### 4.2. Simulated Performance

To illustrate the application of SIR to ESCAT measurements, simulation is initially used. In the simulation, the geometry and response function from actual ESCAT measurements over a small study region in Antarctica are used with synthetic  $\mathcal{A}$  and  $\mathcal{B}$  “truth” images (see Fig. 6) to generate simulated  $\sigma^\circ$  measurements. Monte Carlo noise with the expected ESCAT  $K_p$  is added to the measurements. The synthetic images include a number of features to aid in evaluating the resolution enhancement including various width lines, a pyramid feature and two small, closely spaced squares. The squares are approximately 25 km in size and spaced 25 km apart. We note that the synthetic images are not bandlimited (as is required by reconstruction theory). This enables us to evaluate the effects of attempting to reconstruct a non-bandlimited image.

The result of applying the SIR algorithm is shown in Fig. 6. The pixel resolution used in these images is 4.45 km. The effective resolution is, of course, less than the pixel resolution. The results after 30 and 100 iterations of the SIR algorithm are shown when 6 days and 30 days of data are used. For comparison, nonenhanced ESCAT images with a pixel resolution of approximately 25 km ( $5 \times 4.45$  km) are also shown. To generate the nonenhanced images, all measurements whose center falls within a given pixel are used to estimate  $\mathcal{A}$  and  $\mathcal{B}$  using linear (in dB) regression.

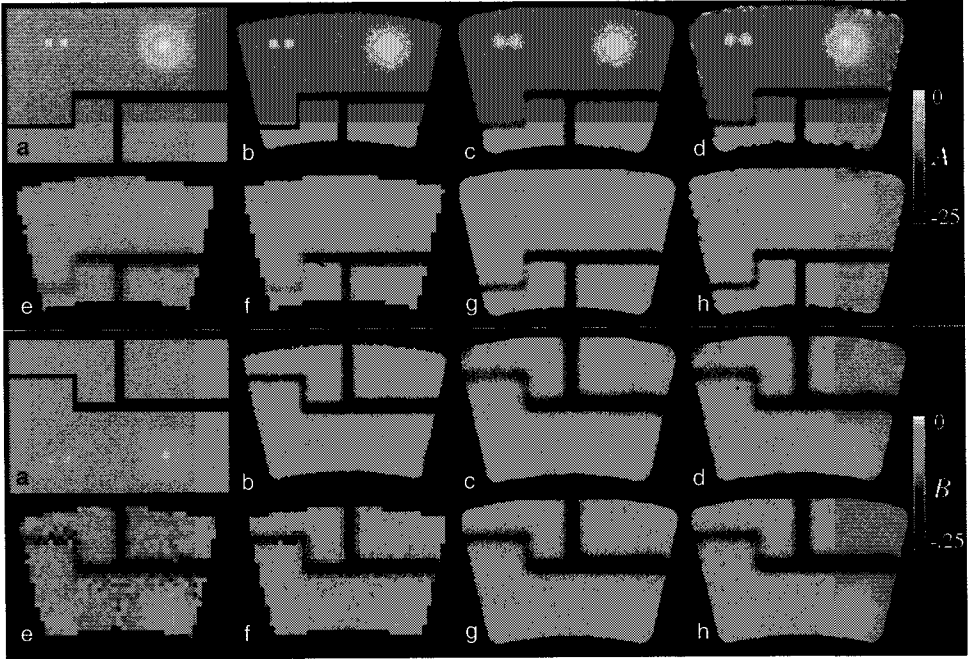


Fig. 6. ESCAT resolution enhancement simulation results. The pixel resolution is 4.45 km. The upper panel contains  $A$  images while the lower panel contains  $B$  images. The subimages in each panel are: a) true synthetic image, b) simulated NSCAT SIR result, c) simulated ESCAT result for 30 iterations with 6 days of data, d) simulated ESCAT result for 1000 iterations with 6 days of data, e) nonenhanced (see text) results for 6 days, f) non-enhanced for 30 days of data, g) simulated ESCAT result for 30 iterations with 30 days of data, and h) simulated ESCAT result for 1000 iterations with 30 days of data.

For ease of display and comparison, each pixel of the nonenhanced images was replicated five times in each direction to expand its size to match the equivalent area of the other images. Also shown are NSCAT comparison images, the result of SIRF processing 6 days of synthetic NSCAT measurements generated in a manner similar to the synthetic ESCAT measurements. The NSCAT images use 50 iterations of SIRF<sup>44</sup>.

Examining Fig. 6 we note that the NSCAT image has better resolution than the ESCAT image. Due to the lower sidelobes of ESCAT data, the improvement in the ESCAT resolution is lower than NSCAT for the same number of iterations. The effective resolution of the ESCAT images improves for both increasing iterations and also when more measurements are incorporated. As more measurements are included, the noise level in the images drops. Also, the SIR-processed images are subjectively better than the nonenhanced images both in terms of resolution and noise level. Careful examination reveals low amplitude artifacts due to low pass filtering of the true image in all of the enhanced resolution images. This effect will be discussed in greater detail later. We note that the  $B$  images exhibit somewhat

lower resolution than the  $\mathcal{A}$  images, an effect previously noted<sup>35</sup>.

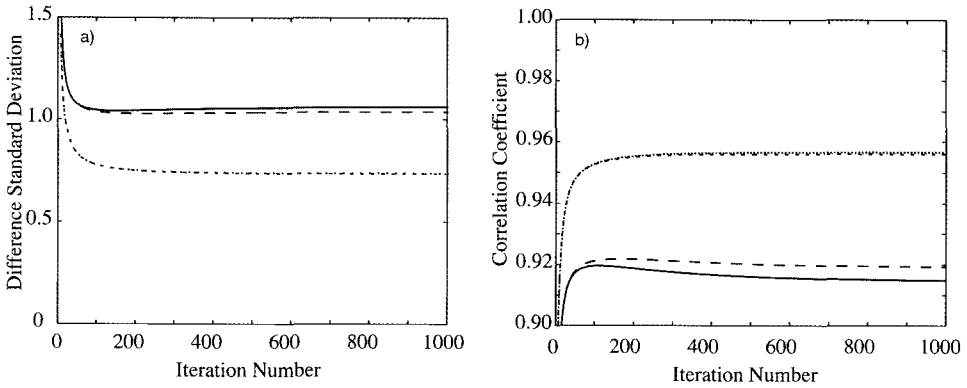


Fig. 7. a) Plot of the standard deviation of the difference between the images versus iteration number. b) Plot of the correlation coefficients of the images version iteration number. In both plots the solid line is the true-ESCAT 6 day image, the long dash line is the true-ESCAT 30 day image, the short dash line is the 6 day ESCAT-NSCAT, and the dotted line is the 30 day ESCAT-NSCAT. In a) the latter two lines lie nearly on top of each other and can not be distinguished in this plot.

To objectively quantify the resolution enhancement, we compare the error between the ESCAT and true images and, to gain insight with later comparisons of actual data, with NSCAT images. As a metric, we compute the standard deviation of the difference of the respective images. We also compute the correlation coefficients. These metrics are computed for the  $\mathcal{A}$  images as a function of the iteration number in Fig. 7. After an initially steep decrease, the ESCAT-true standard deviation of the 6 day measurement set bottoms out at about 75 iterations and begins a slow rise. The ESCAT-true 30 day set is similar, with a much slower rise and a minimum at about 200 iterations. The ESCAT-NSCAT standard deviation for both the 6 day and 30 day cases, look similar, though lower and with a minima at approximately 500 iterations. The correlation coefficient exhibits a behavior consistent with the standard deviation, with a rapid initial increase in correlation, a peak and then a gradual decrease in the correlation. The minima for the ESCAT-true correlation is at approximately 100 and 150 for the 6 day and 30 day cases, respectively. For ESCAT-NSCAT, the peak is at 500 iterations.

Detailed examinations of the images and their spectra at each iteration (not shown) suggest that the initial standard deviation decrease and correlation increase are due primarily to the recovery of the signal in the main lobe of the aperture function response. As the iterations continue, the difference between the reference signal and the image decreases. The image component due to the noise begins small and gradually increases with iteration. Eventually, the increasing noise begins to dominate over the decreasing signal error, leading to the increase in standard deviation and decrease in correlation. Consideration of the sensitivity of the particular



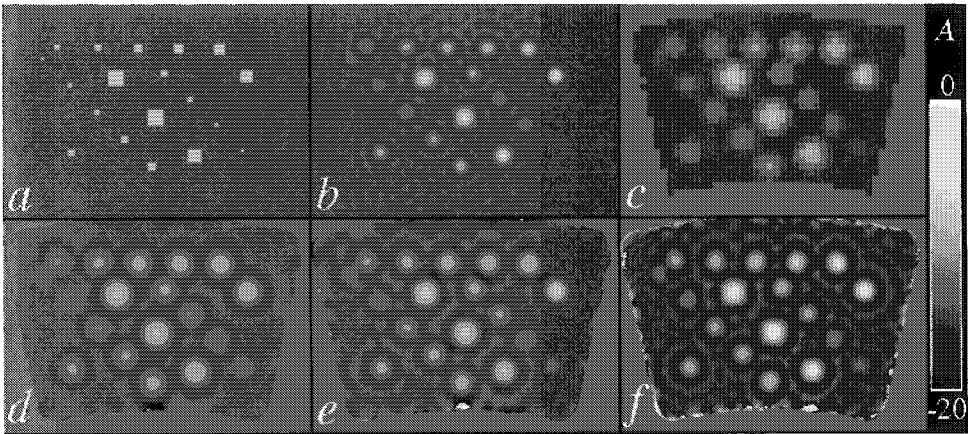


Fig. 8.  $A$  image comparison using six days of data. Actual measurement locations over a small polar region are used to create noisy synthetic  $\sigma^\circ$  measurements of the true image. a) True synthetic  $A$  image containing variously sized targets. b) True image low pass filtered to 25 km resolution. Note ringing artifacts due to the low pass filter. c) Non-enhanced resolution image (see text). d) Result after 30 iterations of SIR. e) Result after 100 iterations of SIR. f) Result after 840 iterations of SIR.

application of the image data dictates the level of noise enhancement that can be tolerated and hence the resolution enhancement. The available computational resources may also be a factor, particularly for large images. While more iteration leads to better resolution, it also increases the noise level which is the primary limitation for ESCAT enhancement.

To better understand the tradeoff between resolution and the number of iterations, the results of a second simulation are shown in Fig. 8. Only the  $A$  images are shown. The simulation procedure is identical to the previous one but with a different synthetic image. Six days of data are used (though as could be expected, better results are obtained with 30 days of data). The synthetic image (Fig. 8a) contains a set of different sized boxes to aid in the evaluation of the resolution of the resulting images. The boxes range in size from 8.9 to 71 km. Along the top row the boxes increase from 17.8 km to 35.6 km. The simulated  $\sigma^\circ$  measurements are generated from this image. A lowpass filtered version of the true image is shown in Fig. 8b. This reflects an ideally reconstructed image using an ideal lowpass filter with a cutoff at 25 km. For reference Fig. 8c shows the corresponding nonenhanced  $A$  image. Figures 8d-f show the results of SIR after 30, 100, and 840 iterations. The bright edge around the SIR images is a simulation artifact. The increasing sharpness of the SIR images with increasing iteration is apparent.

While the larger diameter of the dark rings for the SIR images initially suggests that the resolution is not as good as the lowpass filtered true image, examination of the 3 dB widths of the boxes in the image estimates suggests otherwise. Figure 9 plots the  $A$  values in the images in Fig. 8 along a line through the top row of boxes.

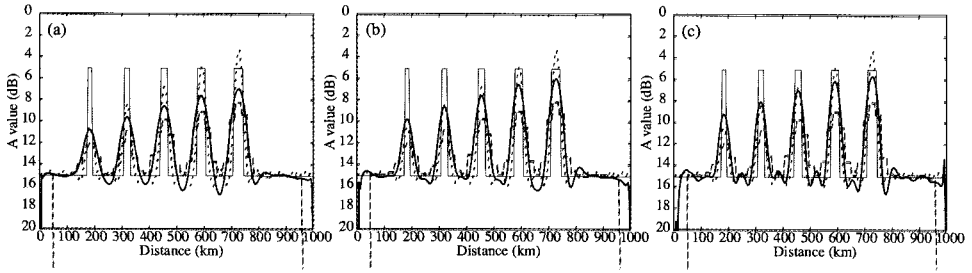


Fig. 9. Plot of  $\mathcal{A}$  values extracted from the images in the previous figure across the top row of spots. The light solid line is the true image. The dark solid line is the SIR estimate. The dotted line is the 25 km low-pass filtered true image. The dashed line is the nonenhanced image data. a) 30 iterations. b) 100 iterations. c) 840 iterations.

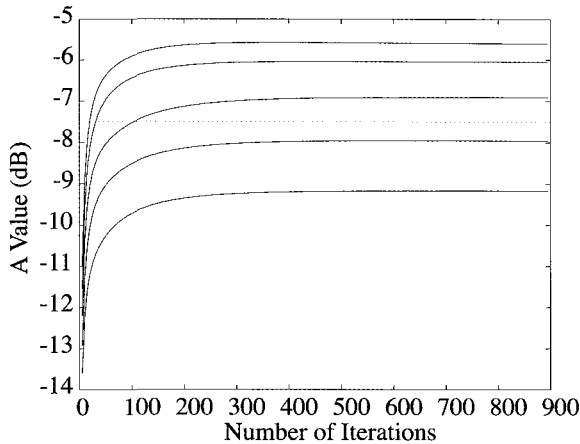


Fig. 10. Plot of the peak  $\mathcal{A}$  value in the SIR image over each of the top row of squares as a function of the number of iterations. The dotted line is at  $-7.5$  dB, the 3 dB point.

Three plots are shown, one for each SIR iteration count considered. In these plots the dark solid line is the SIR image estimate while the dotted line is the lowpass filtered true image. Comparing the three plots, the adaptation of the SIR image with iteration is apparent. Examining the SIR plot and the lowpass filtered plot, it can be seen that the box peaks for SIR are above the lowpass peaks for small objects and somewhat below for large objects. The closest overall match occurs for 100 iterations. Figure 10 plots the heights of each of the boxes as a function of iteration. The dotted line corresponds to the 3 dB point. Thus, using a 3 dB effective resolution criterion, we conclude that the smallest resolvable box for 6 days of data and 100 iterations is approximately 25 km. However, smaller features are more readily apparent (as a larger peak) than predicted by the lowpass filtered image suggesting that there is, in fact, information at higher frequencies than this. Further, as the algorithm is iterated longer, the effective resolution continues to

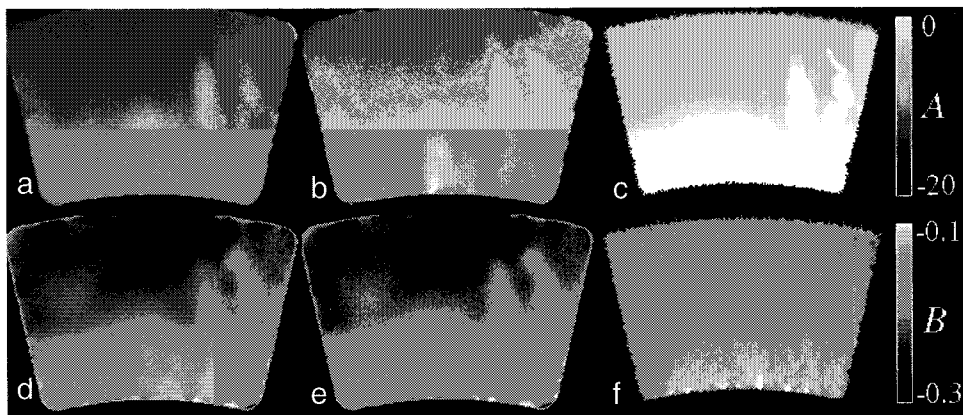


Fig. 11.  $\mathcal{A}$  image resolution enhancement results for actual data over a small study region in Antarctica. The pixel resolution is 4.45 km. a) ESCAT  $\mathcal{A}$  result for 35 iterations, b) ESCAT  $\mathcal{A}$  result for 1000 iterations. c) NSCAT  $\mathcal{A}$  result. d) ESCAT  $\mathcal{B}$  result for 35 iterations, e) ESCAT  $\mathcal{B}$  result for 1000 iterations. f) NSCAT  $\mathcal{B}$  result.

improve, albeit slowly. Based on these simulations, it is possible to obtain enhanced resolution images from ESCAT measurements even for short imaging periods.

### 4.3. Actual Data

Having used simulation to evaluate the resolution enhancement of ESCAT data, actual data is now considered. One of the difficulties with using actual data is that the true values of  $\mathcal{A}$  and  $\mathcal{B}$  are not known, making a quantitative evaluation of the resolution enhancement very difficult. Instead, we compare the enhanced resolution images to data from other sensors. This comparison is complicated by the fact that sensors operate at different frequencies and so the surface response characteristics vary. Nevertheless with this limitation in mind, the correlation between the sensors can provide a measure of the resolution enhancement. Two study regions are considered: polar (Antarctica) and mid-latitude (Amazon).

In the polar region example, 6 days of data are used to generate  $\mathcal{A}$  and  $\mathcal{B}$  images of Wilkes land in Antarctica. The results are shown in Fig. 11 for two different SIR iterations. An NSCAT-derived image of the same location and time is shown. Plots of the ESCAT-NSCAT standard deviation and correlation are shown in Fig. 12. We note that ESCAT and NSCAT operate at different frequencies (5.6 GHz versus 14.0 GHz) which can be expected to have somewhat different responses to surface features. Nevertheless, similar features are observed in the  $\mathcal{A}$  images from both sensors, with the ESCAT images appearing like lowpass versions of the NSCAT images. Greater differences are evident in the  $\mathcal{B}$  images. While this is not well-understood, it may be due to the differences in scattering at different frequencies from interannual layers in the Antarctic firn or azimuth modulation of the backscatter with azimuth angle. Careful examination of the ESCAT  $\mathcal{A}$  images reveals somewhat sharper edges

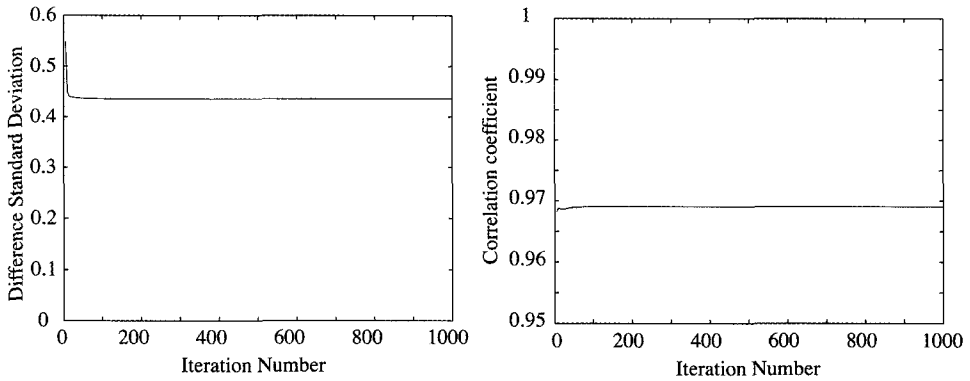


Fig. 12. a) Plot of the standard deviation of the difference between the ESCAT and NSCAT *A* image results versus iteration number. b) Plot of the correlation coefficient between the ESCAT and NSCAT *A* image results versus iteration number.

on the features in the image resulting from more iterations, suggesting higher effective resolution. As in the simulations, the ESCAT-NSCAT standard deviations and correlations show rapid initial improvement, more gradual improvement, a minimum and then very gradual degradation with iteration due to noise enhancement.

For the Amazon region, a small study region covering part of the Amazon river is considered. Non-enhanced and resolution enhanced images from ESCAT, NSCAT, and SeaWinds data are compared in Fig. 13. This image compares the output of both the SIR and AVE algorithms. The AVE algorithm<sup>35</sup> is defined as the first iteration of SIR. In this example, images from both types of SeaWinds measurements are shown. SIR with the actual antenna response was used with SeaWinds 25 km ‘egg measurements’ while SIRF was used on SeaWinds  $6 \times 25$  km ‘slice measurements’. SIRF was also used for NSCAT. For this comparison the number of iterations was limited to 30 for SIR and 50 for SIRF. Again, note that additional iteration improves the resolution, but also increase the noise. However, this comparison restricts the number of iterations to a small number. The imaging period varies from 4 days for SeaWinds to 6 days for ESCAT and NSCAT. The pixel resolution is 4.45 km. Examining these images it is apparent that SIR yields improved resolution images compared to both AVE and non-enhanced. While the resolution improvement for ESCAT is less than the other sensors, ESCAT resolution enhancement is effective.

## 5. Conclusion

This Chapter has considered spatial resolution enhancement by reconstruction from irregularly sampled microwave sensor observations. Given sufficiently dense sampling, an enhanced resolution image of the surface can be generated using reconstruction from the sensor observations. In the resulting image the attenuation resulting from effective aperture function is compensated for, exclusive of the spectral nulls in the effective aperture function. The aperture function arises from the net

ing high frequency information attenuated by the aperture function in oversampled data. The enhancement is limited by nulls in the aperture function and the sampling density. Noise enhancement in the reconstruction can also be a limiting factor. When required, the sample density can be increased by combining data from multiple passes, at the expense of temporal averaging and reduced temporal resolution. The reconstructed images have “enhanced resolution” since the effective resolution can be much finer than the nominal 3 dB sensor resolution. Additive and multiplicative ART can be used as reconstruction algorithms, though the derivative SIR algorithm is more robust in the presence of noise. SIR has been successfully applied to scatterometer and radiometer data<sup>9,29,31,35</sup>.

As an illustration, the technique is applied to ESCAT data, which is a very demanding application due to the windowed aperture function in the data. Tradeoffs for ESCAT are considered and SIR is shown to improve the resolution of the ESCAT data when multiple orbits are combined. ESCAT results are compared to NSCAT and Seawinds.

## References

1. J.L., Alvarez-Perez, S.J. Marshall, and K. Gregson, Resolution Improvement of ERS Scatterometer Data Over Land by Wiener Filtering, *Remote Sens. Environ.*, **71**, pp. 261–271, 2000.
2. E. Attema, The Active Microwave Instrument Onboard the ERS-1 Satellite, *Proc. IEEE*, **79**, 6, pp. 791–799, 1991.
3. B.G. Baldwin, W.J. Emery, and P.B. Cheeseman, Higher Resolution Earth Surface Features from Repeat Moderate Resolution Satellite Imagery, *IEEE Trans. Geosci. Remote Sens.*, **36**, 1, pp. 244–255, 1998.
4. C. Byrne, *et al.*, Image Restoration and Resolution Enhancement, *J. Opt. Soc. Am.*, **73**, 11, pp 1481–1487, 1983.
5. B. Caccin, C. Roberti, P. Russo, and A. Smaldone, The Backus–Gilbert Inversion Method and the Processing of Sampled Data, *IEEE Trans. Signal Process.*, **40**, pp. 2823–2825, 1992.
6. Y. Censor, Finite Series-Expansion Reconstruction Methods, *Proc. IEEE*, **71**, 3, pp. 409–419, 1983.
7. W.B. Davis, *Enhanced Resolution Imaging from Remotely Sensed Microwave Data*, Masters Thesis, Brigham Young University, 1983.
8. J.E.H. Dyal, *TRMM Sigma-0 Vegetation Signatures and Other Studies*, Masters Thesis, Brigham Young University, 1999.
9. D.S. Early and D.G. Long, Image Reconstruction and Enhanced Resolution Imaging From Irregular Samples, *IEEE Trans. Geosci. Remote Sens.*, **39**, 2, pp. 291–302, 2001.
10. D.S. Early and D.G. Long, Error Characteristics of the SIR Resolution Enhancement Algorithm, *Proc. Int. Geosci. Rem. Sens. Sym.*, pp. 124–126, Lincoln, Nebraska, 27–31 May, 1996.
11. D.S. Early and D.G. Long, Azimuth Modulation of C-band Scatterometer sigma-0 Over Southern Ocean Sea Ice, *IEEE Trans. Geosci. Remote Sens.*, **35**, 5, pp. 1201–1209, 1997.
12. T. Elfving, On Some Methods for Entropy Maximization and Matrix Scaling, *Linear Algebra and its Applications*, **34**, pp. 321–339, 1980.

13. M.R. Farrar and E.A. Smith, Spatial Resolution Enhancement of Terrestrial Features Using Deconvolved SSM/I Brightness Temperatures, *IEEE Trans. Geosci. Remote Sens.*, **30**, 2, pp 349–355, 1992.
14. P. Gagne and H.H. Arsenault, Using Information from Multiple Low-Resolution Images to Increase Resolution, *Proc. of SPIE*, **1564**, pp 656–663, 1991.
15. K. Gröchenig, Reconstruction Algorithms in Irregular Sampling, *Mathematics of Computation*, **59**, 199, pp. 181–194, 1992.
16. N.C. Grody, Classification of Snow Cover and Precipitation Using the Special Sensor Microwave/Imager, *J. Geophys. Res.*, **94**, pp. 7423–7435, 1991.
17. P. Gilbert, Iterative Methods for the Three-Dimensional Reconstruction of an Object from Projections, *J. Theor. Bio.*, **36**, pp. 105–117, 1972.
18. J.P. Hollinger, J.L. Pierce, G.A. and Poe, SSM/I Instrument Evaluation, *IEEE Trans. Geosci. Remote Sensing*, **28**, 5, pp. 781–790, 1990.
19. G. Hyland and N. Young, Wind-induced directional anisotropy of microwave backscatter and its impact on imaging of the Antarctic continental snow cover, *Proc. Int. Geosci. Rem. Sens. Sym.*, pp. 1988–1990, 1998.
20. T.J. Jackson and T.J. Schmugge, Algorithm for the Passive Microwave Remote Sensing of Soil Moisture, in *Microwave Radiometry and Remote Sensing Applications* (P. Pampaloni, ed.), (Zeist), pp. 3–17, VSP, 1989.
21. A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ, 1989.
22. A.S. Jones and T.H.V. Haar, Passive Microwave Sensing of Cloud Liquid Water Over Land Regions, *J. Geophys. Res.*, **95**, pp. 16673–16683, 1990.
23. W.L. Jones, L.C. Schroeder, D.H. Boggs, E.M. Bracalente, R.A. Brown, G.J. Dome, W.J. Pierson, and F.J. Wentz. The SEASAT-A Satellite Scatterometer: The Geophysical Evaluation of Remotely Sensed Wind Vectors Over the Ocean. *J. Geophys. Res.*, **87**, C5, 3297–3317, 1982.
24. T. Kawanishi, H. Takamatsu, T. Kosu, K. Okamoto, H. and Kumagai, TRMM Precipitation Radar, *Proc. IGARSS'93*, pp. 423–425, Tokyo, Japan, Aug., 1993.
25. R.L. Lagendijk and J. Biemond, *Iterative Identification and Restoration of Images*, Kluwer Academic Publishers, Boston, 1991.
26. P. Lecomte, A. Cavanie, and F. Gohin, Recognition of Sea Ice Zones using ERS-1 Scatterometer Data, *Proc. Int. Geosci. Rem. Sens. Sym.*, pp. 855–857, 1993.
27. M. Ledroit, F. Remy, and J.F. Minster, Observation of the Antarctic ice sheet with the Seasat Scatterometer: relation to katabatic wind intensity and direction, *J. Glaciology*, **39**, 132, pp. 385–396, 1993.
28. D.G. Long and M.R. Drinkwater, Azimuth Variation in Microwave Scatterometer and Radiometer Data Over Antarctica, *IEEE Trans. Geosci. Remote Sens.*, **38**, 4, pp. 1857–1870, 2000.
29. D.G. Long and M.R. Drinkwater, Cryosphere Applications of NSCAT Data, *IEEE Trans. Geosci. Remote Sens.*, Vol. 37, 3, pp. 1671–1684, 1999.
30. D.G. Long, Comparison of TRMM and NSCAT Observations of Surface Backscatter Over the Amazon Rainforest, *Proc. Int. Geosci. Rem. Sens. Sym.*, pp. 1879–1881, Seattle, Washington, 6–10 July, 1998.
31. D.G. Long and D. Daum, Spatial Resolution Enhancement of SSM/I Data, *IEEE Trans. Geosci. Remote Sens.*, **36**, pp. 407–417, 1997.
32. D.G. Long and G.B. Gkouson, Calibration of Spaceborne Scatterometers Using Tropical Rainforests, *IEEE Trans. Geosci. Remote Sens.*, **34**, 2, pp. 413–424, 1996.
33. D.G. Long and P. Hardin, Vegetation Studies of the Amazon Basin Using Enhanced Resolution Seasat Scatterometer Data, *IEEE Trans. Geosci. Remote Sens.*, **32**, 2,

- pp. 449–460, 1994.
34. D.G. Long and M.R. Drinkwater, Greenland Observed at High Resolution by the Seasat-A Scatterometer, *J. Glaciology*, **32**, 2, pp. 213–230, 1994.
  35. D. Long, P. Hardin, and P. Whiting, Resolution Enhancement of Spaceborne Scatterometer Data, *IEEE Trans. Geosci. Remote Sens.*, **31**, pp. 700–715, 1993.
  36. D.G. Long and J.M. Mendel, Identifiability in Wind Estimation from Wind Scatterometer Measurements, *IEEE Trans. Geosci. Remote Sensing*, **29**, 2, pp. 268–276, 1991.
  37. M.J. McFarland, R.L. Miller, and C.M.U. Neale, Land Surface Temperature Derived From the SSM/I Passive Microwave Brightness Temperatures, *IEEE Trans. Geosci. Remote Sensing*, Vol. 28, pp. 839–845, 1990.
  38. F. Naderi, M.H. Freilich, and D.G. Long, Spaceborne Radar Measurement of Wind Velocity Over the Ocean—An Overview of the NSCAT Scatterometer System, *Proc. IEEE*, **79**, 6, pp. 850–866, 1990.
  39. E.G. Njoku, J.M. Stacey, and F.T. Barath, The Seasat Scanning Multichannel Microwave Radiometer (SMMR): Instrument Description and Performance, *IEEE J. Ocean. Eng.*, **EO-5**, No. 2, pp. 100–115, 1980.
  40. W.S. Olson, C-L Yeh, J.A. Weinman, R.T. Chin, Resolution Enhancement of Multichannel Microwave Imagery from the Nimbus-7 SMMR for Maritime Rainfall Analysis, *J. Atm. and Oceanic Tech.*, **3**, pp 422–432, 1986.
  41. A.V. Oppenheim and A.S. Willsky, *Signals and Systems*, Prentice Hall, Englewood Cliffs, NJ, 1983.
  42. P. Pampaloni and S. Paloscia, Microwave Emission and Plant Water Content: A Comparison Between Field Measurement and Theory, *IEEE Trans. Geosci. Remote Sensing*, **24**, pp. 900–904, 1986.
  43. G.A. Poe, Optimum Interpolation of Imaging Microwave Radiometer Data, *IEEE Trans. Geosci. Rem. Sens.*, **GE-28**, pp 800–810, 1990.
  44. Q.P. Remund, D.E. Early, and D.G. Long, Optimization of SIRF for NSCAT, Microwave Earth Remote Sensing Laboratory internal report, Brigham Young University, Provo, UT, MERS#96-003, 3 July, 1999. (Available via <http://www.mers.byu.edu/Papers/reports>).
  45. F. Remy, M. Ledroit, and J.F. Minster, Katabatic Wind Intensity and Direction over Antarctica Derived from Scatterometer Data, *Geophys. Res. Letters*, **19**, pp. 1021–1024, 1992.
  46. W.D. Robinson, C. Kummerow, and W.S. Olson, A Technique for Enhancing and Matching the Resolution of Microwave Measurements from the SSM/I Instrument, *IEEE Trans. Geosci. Rem. Sens.*, **30**, 3, pp. 419–429, 1992.
  47. R. Sethmann, B.A. Burns, and G.C. Heygster, Spatial Resolution Improvement of SSM/I Data with Image Restoration Techniques, *IEEE Trans. Geosci. Rem. Sens.*, **32**, 6, pp. 1144–1151, 1994.
  48. B. Sharpe and A. Kerr, Multichannel Fusion Techniques for Resolution Enhancement During Ground Processing, *Proc. Int. Geosci. Rem. Sens. Sym.*, Espoo, Finland, pp. 1495–1462, 1991.
  49. K.V. Shettigara, A Linear Transformation Technique for Spatial Enhancement of Multispectral Images Using a Higher Resolution Data Set, *Proc. Int. Geosci. Rem. Sens. Sym.*, pp. 2615–2618, 1989.
  50. M.W. Spencer, C. Wu, and D.G. Long, Improved Resolution Backscatter Measurements with the SeaWinds Pencil-Beam Scatterometer, *IEEE Trans. Geosci. Remote Sens.*, **38**, 1, pp.89–104, 2000.
  51. H. Stephen and D.G. Long, Multi-spectral Analysis of the Amazon Basin using Sea-

- Winds, ERS, Seasat Scatterometers, TRMM-PR and SSM/I, *Proc. Int. Geosci. Rem. Sens. Sym.*, Toronto, Canada, 24–28 June, pp. 1780–1782, 2002.
52. P. J. Stephen and A.S. Jones, Computationally Efficient Discrete Backus–Gilbert Footprint-Matching Algorithm, *IEEE Trans. Geosci. Rem. Sens.*, **40**, 8, pp. 1865–1878, 2002.
  53. A. Stogryn, Estimates of Brightness Temperatures from Scanning Radiometer Data, *IEEE Trans. Antennas Propagat.*, **AP-26**, 5, pp 720–726, 1978.
  54. R.H. Thomas, R.A. Bindschadler, R.L. Cameron, F.D. Carsey, B. Holt, T.J. Hughes, C.W.M. Swithinbank, I.M. Whillans, and H.J. Zwally, Satellite Remote Sensing for Ice Sheet Research, *NASA Technical Memorandum*, 86233, 27, 1985.
  55. F.T. Ulaby, R.K. Moore, and A.K. Fung, *Microwave Remote Sensing — Active and Passive*, Vols. 1 and 2, Addison–Wesley Publishing Co., Reading, Mass., 1981.
  56. N.P. Walker, Limitations of the possible resolution enhancement of ERS-1 scatterometer images, *IEEE Trans. Geosci. Rem. Sens.*, **35**, 1, pp. 196–198, 1997.
  57. F.J. Wentz, Measurement of Oceanic Wind Vector Using Satellite Microwave Radiometers, *IEEE Trans. Geosci. Rem. Sens.*, **30**, 5, pp. 960–972, 1992.
  58. M.E. Willis, *Algebraic Reconstruction Algorithms for Remote Sensing Image Enhancement*, Master’s Thesis, Brigham Young University, Provo, Utah, 2000.
  59. V.R. Wismann, K. Boehnke, and C. Schmullius, Monitoring Ecological Dynamics in Africa with the ERS-1 Scatterometer, *Proc. Int. Geosci. Rem. Sens. Sym.*, pp. 1523–1525, 1995.



This page is intentionally left blank

# **Statistical Pattern Recognition**

This page is intentionally left blank

## CHAPTER 12

### ADVANCED CLASSIFICATION TECHNIQUES: PARTIALLY SUPERVISED APPROACHES

Lorenzo Bruzzone

*Department of Information and Communication Technology  
Via Sommarive 14, I-38050, Trento, Italy  
E-mail: lorenzo.bruzzone@ing.unitn.it*

Roberto Cossu

*Department of Information and Communication Technology  
Via Sommarive 14, I-38050, Trento, Italy  
E-mail: roberto.cossu@ing.unitn.it*

Advanced classification techniques for a regular updating of land-cover maps are proposed that are based on the use of multitemporal remote-sensing images. Such techniques are developed within the framework of partially supervised approaches, which are able to address the updating problem under the realistic but critical constraint that, for the image to be classified (i.e., the most recent of the considered multitemporal data set), no ground truth information is available. Two different approaches are considered. The first approach is based on an independent analysis of the information contained in each single image of the considered multitemporal series; the second approach exploits the temporal correlation between pairs of images acquired at different times in the classification process. In the context of such approaches, both parametric and non-parametric classifiers are considered. In addition, in order to design a reliable and accurate classification system, multiple classifier architectures composed of partially supervised algorithms are investigated. Experimental results obtained on a real multitemporal data set confirm the effectiveness of the proposed approaches.

#### 1. Introduction

One of the major problems in geographical information systems (GISs) consists in defining strategies and procedures for a regular updating of land-cover maps stored in the system databases. This crucial task can be carried out by using remote-sensing images regularly acquired by space-born sensors in the specific investigated areas. Such images can be analyzed with automatic classification techniques in order to derive updated land-cover maps. The classification process can be performed by considering either the information contained in a single image<sup>1</sup> or the information contained in a multitemporal series of images of the same area<sup>2,3,4</sup> (i.e., by exploit-

ing the temporal correlation between images acquired at different times). The latter approach is called “cascade classification” and allows one to increase the categorization accuracy. However, at the operating level, both aforementioned approaches are usually based on supervised classification algorithms. Consequently, they require the availability of ground truth information for the training of the classifiers. Unfortunately, in many real cases, it is not possible to rely on training data for all the images necessary to ensure an updating of land-cover maps that is as frequent as required by applications. This prevents all the remotely sensed images acquired in the investigated area from being used to update land-cover maps. For these reasons, the process of temporal updating of land-cover maps results in a complex and challenging problem.

In this chapter, we present two different classification approaches that overcome the above-described drawback of land-cover monitoring systems. These approaches are based on novel partially supervised classification techniques that permit the classifier to generate an accurate land-cover map from the new image even when the related training set is not available. The first approach we present performs the classification process by considering independently the information contained in each single image of the available multitemporal series, whereas the second approach is developed within the framework of cascade classification.

It is worth noting that, given the intrinsic complexity of the problem addressed, the aforementioned approaches may result in classifiers that are less reliable and less accurate than the corresponding supervised ones. In order to overcome this drawback, in this chapter, we also investigate the effectiveness of the use of a multiple classifier system (MCS) composed of partially supervised classifiers<sup>4</sup>. The considered MCS is made up of both parametric and non-parametric classifiers neural-network classification approaches. It is worth noting that the use of non-parametric classification approaches in the MCS allows one to analyze multisensor and/or multisource remote-sensing images, which often play a fundamental role in complex classification problems.

Experimental results obtained on a multitemporal data set related to the Island of Sardinia (Italy) confirm the effectiveness of the proposed approaches in the framework of both independent and cascade classification.

The chapter is organized into seven sections. Section 2 reports the formulation of the problem. Section 3 and Section 4 present the partially supervised classification problem in the framework of the independent-classification and cascade-classification approaches, respectively. In Section 5 an MCS composed of partially supervised classifiers is described. Experimental results are reported in Section 6. Finally, in Section 7, discussion is provided and conclusions are drawn.

## 2. General Formulation of the Problem

Let us consider a classifier for the periodical monitoring of a specific geographical area. Let  $\mathbf{X}_1 = \{x_1^1, x_2^1, \dots, x_B^1\}$  and  $\mathbf{X}_2 = \{x_1^2, x_2^2, \dots, x_B^2\}$  denote two multispectral

images composed of  $B$  pixels and acquired in the area under analysis at the times  $t_1$  and  $t_2$ , respectively. Let  $x_j^1$  and  $x_j^2$  be the  $1 \times d$  feature vectors associated with the  $j$ -th pixels of the images (where  $d$  is the dimensionality of the input space), and  $\Omega = \{\omega_1, \omega_2, \dots, \omega_C\}$  be the set of  $C$  land-cover classes that characterize the geographical area considered at both  $t_1$  and  $t_2$ . Let  $l_j^1$  and  $l_j^2$  be the classification label of the  $j$ -th pixel at the time  $t_1$  and  $t_2$ , respectively. Finally, let  $X_1$  and  $X_2$  be two multivariate random variables representing the pixel values (i.e., the feature vector values) in  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , respectively.

In the formulation of the proposed approach, we make the following assumptions:

- the same set of  $C$  land-cover classes (i.e.,  $\Omega$ ) characterize the area considered over time (only the spatial and spectral distributions of such classes are supposed to vary);
- a reliable training set  $\mathbf{Y}_1$  for the image  $\mathbf{X}_1$  acquired at  $t_1$  is available;
- a training set  $\mathbf{Y}_2$  for the image  $\mathbf{X}_2$  acquired at  $t_2$  is not available.

It is worth noting that the first assumption, even if not verified in all possible applications, is reasonable in a wide range of real problems. Examples of such applications include studies on forestry, territorial management, and natural-resource monitoring on a national or even continental scale<sup>7,8,9</sup>.

In the aforementioned hypotheses, the proposed system aims at performing a robust and accurate classification of  $\mathbf{X}_2$  by exploiting the image  $\mathbf{X}_1$ , the training set  $\mathbf{Y}_1$ , and the image  $\mathbf{X}_2$ .

### 3. Partially Supervised Independent-Classification Approach

In the context of the Bayes decision theory, the decision rule adopted by a classifier is expressed as follows<sup>1,10</sup>:

$$l_j^1 = \omega_m \in \Omega \text{ if and only if } P_1(\omega_m|x_j^1) = \max_{\omega_h \in \Omega} \{P_1(\omega_h|x_j^1)\} \quad (1)$$

where  $P_1(\omega_i|x_j^1)$  is the value of the probability that the  $j$ -th pixel of the image  $\mathbf{X}_1$  belong to the class  $\omega_i$  given the observation  $x_j^1$ . It is worth noting that the subscript 1 is used here to stress the dependences of statistical terms on the considered image  $\mathbf{X}_1$ . Equation 1 can be rewritten as<sup>1,10</sup>:

$$l_j^1 = \omega_m \in \Omega \text{ if and only if } P_1(\omega_m)p_1(x_j^1|\omega_m) = \max_{\omega_h \in \Omega} \{P_1(\omega_h)p_1(x_j^1|\omega_h)\} \quad (2)$$

where  $P_1(\omega_i)$  is the *a priori* probability of the class  $\omega_i$  in the image  $\mathbf{X}_1$  and  $p_1(x_j^1|\omega_i)$  is the value of the conditional density function for the pixel  $x_j^1$ , given the class  $\omega_i \in \Omega$ . According to (2), the training phase of the classifier consists in the estimations of the *a priori* probability  $P_1(\omega_i)$  and the conditional density  $p_1(X_1|\omega_i)$  for each class  $\omega_i \in \Omega$ . Such estimates can be obtained by using classical supervised approaches, which exploit the information that is present in the considered training set  $\mathbf{Y}_1$ <sup>1,10</sup>.

Let us now assume that, at the time  $t_2$ , a new land-cover map of the study area is required. Let us also assume that at  $t_2$  the corresponding training set is not available. This prevents the generation of the required land-cover map as the training of the classifier cannot be performed (i.e., both the *a priori* probability  $P_2(\omega_i)$  and the conditional density function  $p_2(X_2|\omega_i)$  of each class  $\omega_i \in \Omega$  in the new image  $\mathbf{X}_2$  cannot be estimated by traditional supervised approaches). At the same time, it is not possible to apply the classifier trained on the image  $\mathbf{X}_1$  to the image  $\mathbf{X}_2$  because, in general, the estimates of statistical class parameters at  $t_1$  do not provide accurate approximations for the same terms at  $t_2$ . This is due to several factors (e.g., differences in the atmospheric and light conditions at the image-acquisition dates, sensor nonlinearities, different levels of soil moisture, etc.) that alter the spectral signatures of land-cover classes in different images and consequently the distributions of such classes in the feature space.

In this context, we propose an unsupervised retraining technique to derive, for each class  $\omega_i \in \Omega$ , reliable estimates of both  $P_2(\omega_i)$  and  $p_2(X_2|\omega_i)$ , starting from the current classifier parameters obtained in a supervised way at the time  $t_1$ .

The main idea of the proposed technique is that the first approximate estimates of the parameter values that characterize the classes considered at the time  $t_2$  can be obtained by exploiting the classifier parameters estimated at the time  $t_1$  by supervised learning. In particular, for each class  $\omega_i \in \Omega$ , the initial values of both the prior probability  $P_2^0(\omega_i)$  and the conditional density function  $p_2^0(X_2|\omega_i)$  can be approximated by:

$$P_2^0(\omega_i) = \hat{P}_1(\omega_i); \quad p_2^0(X_2|\omega_i) = \hat{p}_1(X_1|\omega_i). \quad (3)$$

where  $\hat{P}_1(\omega_i)$  and  $\hat{p}_1(X_1|\omega_i)$  are the prior probability and the conditional density function, respectively, estimated for the class  $\omega_i$  at the time  $t_1$  by supervised learning. As already pointed out, generally such first estimates do not provide accurate approximations for the statistical parameters of the classes at  $t_2$ . Therefore, we suggest improving such rough estimates by exploiting the information associated with the distribution  $p_2(X_2)$  of the new image  $\mathbf{X}_2$ . In particular, the proposed method is based on the observation that the statistical distribution of the pixel values in  $\mathbf{X}_2$  can be described by a mixed density distribution with as many components as the classes to be recognized:

$$p_2(X_2) = \sum_{i=1}^C P_2(\omega_i) p_2(X_2|\omega_i), \quad (4)$$

where the mixing parameters and the component densities are the *a priori* probabilities and the conditional density functions of the classes, respectively<sup>11</sup>. In this context, the retraining of the Bayesian classifier at the time  $t_2$  becomes a mixture density estimation problem. In our case, this problem involves the estimation of the

parameter vector  $\theta = [\theta_1^2, P_2(\omega_1), \theta_2^2, P_2(\omega_2), \dots, \theta_C^2, P_2(\omega_C)]$ , where each component  $\theta_i^2$  represents the vector of parameters that characterizes the density function  $p_2(X_2|\omega_i)$ , which, given its dependence on  $\theta_i^2$ , can be rewritten as  $p_2(X_2|\omega_i, \theta_i^2)$ . The components of  $\theta$  can be estimated by maximizing a log-likelihood function  $\ell(\mathbf{X}_2|\theta)$  defined as:

$$\ell(\mathbf{X}_2|\theta) = \sum_{j=1}^B \log \left\{ \sum_{i=1}^C P_2(\omega_i|\theta) p_2(x_j^2|\omega_i, \theta) \right\}. \quad (5)$$

The EM algorithm<sup>12,13,14</sup> is one of the most powerful solutions to this type of problem. It consists of two main steps: an expectation step and a maximization step. Both steps are iterated so that, at each iteration, the estimated parameters provide an increase in the log-likelihood function  $\ell(\mathbf{X}_2|\theta)$  until a local maximum is reached. An important aspect of the EM algorithm concerns its convergence properties. Even though convergence can be ensured, it is impossible to guarantee that the algorithm will converge to the global maximum of the log-likelihood function (only in few specific cases is it possible to ensure the convergence to the global maximum). A detailed description of the EM algorithm and its theoretical aspects is beyond the scope of this chapter. We refer the reader to the literature for a more detailed analysis of such an algorithm and its properties<sup>12,14</sup>. In the following subsection, we describe in greater detail the partially supervised independent classification approach for a parametric classification technique.

### 3.1. Independent Normal Bayesian Classifier

To further explain the proposed approach, let us consider, for simplicity, the case in which all classes included in  $\Omega$  can be described by Gaussian distributions. Under this common assumption (widely adopted for multispectral image classification problems), the density function associated with each class  $\omega_r$  at  $t_i$  ( $i = 1, 2$ ) can be completely described by the mean vector  $\mu_r^i$  and the covariance matrix  $\Sigma_r^i$ . Therefore, the vector of parameters to be estimated becomes:

$$\theta = [\mu_1^2, \Sigma_1^2, P_2(\omega_1), \dots, \mu_C^2, \Sigma_C^2, P_2(\omega_C)]. \quad (6)$$

It can be proven that the equations for estimating the statistical terms associated with a generic class  $\omega_r$  are the following<sup>12,13,14</sup> :

$$P_2^{t+1}(\omega_r) = \frac{1}{B} \sum_{j=1}^B \frac{P_2^t(\omega_r) p_2^t(x_j^2|\omega_r)}{p^t(x_j^2)} \quad (7)$$



$$[\mu_r^2]^{t+1} = \frac{\sum_{j=1}^B \frac{P_2^t(\omega_r) p_2^t(x_j^2 | \omega_r)}{p_2^t(x_j^2)} x_j^2}{\sum_{j=1}^B \frac{P_2^t(\omega_r) p_2^t(x_j^2 | \omega_r)}{p_2^t(x_j^2)}} \quad (8)$$

$$[\Sigma_r^2]^{t+1} = \frac{\sum_{j=1}^B \frac{P_2^t(\omega_r) p_2^t(x_j^2 | \omega_r)}{p_2^t(x_j^2)} \{x_j^2 - [\mu_r^2]^{t+1}\}^T \{x_j^2 - [\mu_r^2]^{t+1}\}}{\sum_{j=1}^B \frac{P_2^t(\omega_r) p_2^t(x_j^2 | \omega_r)}{p_2^t(x_j^2)}} \quad (9)$$

where the superscripts  $t$  and  $t + 1$  refer to the values of the parameters at the current and next iterations, respectively, and the superscript  $T$  refers to the vector transpose operation. The estimates are obtained starting from the initial values of the considered parameters and iterating the above equations up to convergence.

The estimates obtained for each class  $\omega_r \in \Omega$  at convergence are the new parameters of the Normal Bayesian classifier at the time  $t_2$ .

At this point, a land-cover map of the analyzed area at the time  $t_2$  can be generated by labeling each pixel  $x_j^2$  in accordance with the Bayesian decision rule:

$$l_j^2 = \omega_m \in \Omega \text{ if and only if } P_2(\omega_m) p_2(x_j^2 | \omega_m) = \max_{\omega_h \in \Omega} \{P_2(\omega_h) p_2(x_j^2 | \omega_h)\} \quad (10)$$

#### 4. Partially Supervised Cascade-Classification Approach

In this section, we propose an extension of the approach described in Section 3. In particular, the methodology described in this section, unlike the previous one, makes use of a cascade-classifier approach<sup>2,3,4</sup> to the categorization of multitemporal remote-sensing images, thus allowing the exploitation of the temporal correlation between successive scenes.

The standard supervised cascade-classifier approach (proposed by Swain<sup>2</sup>) exploits the correlation between multitemporal images in order to increase the classification accuracy in the cases in which training data are available for all the images considered. In our method, we extend the application of the standard supervised cascade-classifier approach to partially supervised classification problems. In particular, we exploit the temporal dependence between land-cover classes to increase the reliability and the accuracy of the partially supervised estimation of the parameters related to the image  $\mathbf{X}_2$ .

The cascade-classifier decision strategy associates a generic pixel  $x_j^2$  of the image  $\mathbf{X}_2$  with a land-cover class according to the following decision rule<sup>2</sup>:

$$l_j^2 = \omega_m \in \Omega \text{ if and only if } P(\omega_m | x_j^1, x_j^2) = \max_{\omega_h \in \Omega} \{P(\omega_h | x_j^1, x_j^2)\} \quad (11)$$

where  $P(\omega_h|x_j^1, x_j^2)$  is the value of the probability that the  $j$ -th pixel of the image belongs to the class  $\omega_h$  at  $t_2$ , given the observations  $x_j^1$  and  $x_j^2$ . Under the conventional assumption of class-conditional independence<sup>2,20</sup>, the decision rule (11) can be rewritten as<sup>3</sup>:

$$l_j^2 = \omega_m \in \Omega \text{ if and only if} \quad (12)$$

$$\sum_{n=1}^C p(x_j^1|\omega_n)p(x_j^2|\omega_m)P(\omega_n, \omega_m) = \max_{\omega_h \in \Omega} \left\{ \sum_{n=1}^C p(x_j^1|\omega_n)p(x_j^2|\omega_h)P(\omega_n, \omega_h) \right\}$$

where  $p(x_j^i|\omega_r)$  is the value of the conditional density function for the pixel  $x_j^i$ , given the class  $\omega_r \in \Omega$ , and  $P(\omega_n, \omega_h)$  is the prior joint probability of the pair of classes  $(\omega_n, \omega_h)$ . The latter term takes into account the temporal correlation between the two images.

We propose to integrate the partially supervised classification of the image  $\mathbf{X}_2$  in the context of the above-described classification rule. As the training set  $\mathbf{Y}_2$  is not available, the density functions of the classes at the time  $t_1$  (i.e.,  $p(X_1|\omega_n)$ ,  $\omega_n \in \Omega$ ) are the only statistical terms of (12) that we can estimate in a completely supervised way. This means that, in order to accomplish the classification task, we should estimate both the density functions of the classes at  $t_2$  ( $p(X_2|\omega_h)$ ,  $\omega_h \in \Omega$ ) and the prior joint probabilities of the classes ( $P(\omega_n, \omega_h)$ ,  $\omega_n \in \Omega$ ,  $\omega_h \in \Omega$ ) in an unsupervised way. It is worth noting that usually the estimation of  $p(X_i|\omega_r)$  ( $\omega_r \in \Omega$ ,  $i=1,2$ ) involves the computation of a parameter vector. The number and nature of the vector components depend on the specific classifier used. Consequently, the procedure to be adopted to accomplish the unsupervised estimation process depends on the technique used to carry out the cascade classification, in particular, on the vector of parameters required by the classifier. In the following subsection, we describe the procedure for the estimation of the above terms for a Normal-Bayesian cascade classifier.

#### 4.1. Normal-Bayesian Cascade Classifier

Let us assume that the probability density function of the generic class  $\omega_r$  at the time  $t_i$  (i.e.,  $p(X_i|\omega_r)$ ,  $\omega_r \in \Omega$ ,  $i=1,2$ ) can be described by a Gaussian distribution (i.e., by a mean vector  $\mu_r^i$  and a covariance matrix  $\Sigma_r^i$ ). Under this common assumption, the mean vectors and the covariance matrices that characterize the conditional density functions of the classes at  $t_1$  can be easily computed by a standard procedure using the training set  $\mathbf{Y}_1$ . Concerning the parameter vector  $\vartheta$  of the classifier to be estimated in a partially supervised way, it consists of the following components:

$$\vartheta = [\mu_1^2, \Sigma_1^2, P(\omega_1, \omega_1), \dots, \mu_C^2, \Sigma_C^2, P(\omega_C, \omega_C)] \quad (13)$$

where the superscript "2" denotes the parameters of the conditional density functions of the classes at the time  $t_2$ . To carry out the partially supervised estimation

process, we propose to adopt a procedure based on the observation that, under the assumption of class-conditional independence over time, the joint density function of the images  $\mathbf{X}_1$  and  $\mathbf{X}_2$  (i.e.,  $p(X_1, X_2)$ ) can be described as a mixture density with  $C \times C$  components (i.e., as many components as possible pairs of classes):

$$p(X_1, X_2) \cong \sum_{n=1}^C \sum_{h=1}^C p(X_1|\omega_n)p(X_2|\omega_h)P(\omega_n, \omega_h). \quad (14)$$

In this context, the estimation of the above terms becomes a mixture-density estimation problem, which can be solved via the EM algorithm<sup>12,13,14,18</sup>. By applying such an algorithm, we can derive the following iterative equations to estimate the components of the vector  $\vartheta$  necessary to accomplish the cascade-classification process<sup>3</sup>:

$$[\mu_h^2]^{t+1} = \frac{\sum_{j=1}^B \sum_{n=1}^C P^t(\omega_n, \omega_h|x_j^1, x_j^2)x_j^2}{\sum_{j=1}^B \sum_{n=1}^C P^t(\omega_n, \omega_h|x_j^1, x_j^2)} \quad (15)$$

$$[\Sigma_h^2]^{t+1} = \frac{\sum_{j=1}^B \sum_{n=1}^C P^t(\omega_n, \omega_h|x_j^1, x_j^2)(x_j^2 - [\mu_h^2]^{t+1})^T(x_j^2 - [\mu_h^2]^{t+1})}{\sum_{j=1}^B \sum_{n=1}^C P^t(\omega_n, \omega_h|x_j^1, x_j^2)} \quad (16)$$

$$P^{t+1}(\omega_n, \omega_h) = \frac{1}{B} \sum_{j=1}^B P^t(\omega_n, \omega_h|x_j^1, x_j^2) \quad (17)$$

where the superscripts  $t$  and  $t+1$  refer to the values of the parameters at the current and next iterations, respectively, the superscript  $T$  refers to the vector transpose operation, and the joint posterior probabilities of the classes are approximated by:

$$P^t(\omega_n, \omega_h|x_j^1, x_j^2) \cong \frac{p(x_j^1|\omega_n)p^t(x_j^2|\omega_h)P^t(\omega_n, \omega_h)}{\sum_{g=1}^C \sum_{f=1}^C p(x_j^1|\omega_g)p^t(x_j^2|\omega_f)P^t(\omega_g, \omega_f)}. \quad (18)$$

It is worth noting that all the previous equations implicitly depend on  $\vartheta$ . Concerning the initialization of the components of the vector  $\vartheta$ , the initial values of the parameters of the density functions of classes at  $t_2$  are obtained by considering the corresponding values estimated at time  $t_1$  by supervised learning, whereas all the prior joint probabilities of classes are assumed to have the same values. It is possible to prove that, at each iteration, the estimated parameters evolve from their

initial values to the final ones by maximizing the following log-likelihood function (the convergence to a local maximum can be proven)<sup>12</sup>:

$$\ell(X_1, X_2|\vartheta) = \sum_{j=1}^B \log \sum_{n=1}^C \sum_{h=1}^C p(x_j^1|\omega_n)p(x_j^2|\omega_h)P(\omega_n, \omega_h). \quad (19)$$

The estimates of the parameters obtained at convergence and those achieved by the classical supervised procedure at the time  $t_1$  are then substituted into (12) in order to accomplish the Normal Bayesian cascade-classification process. We refer the reader to Bruzzone and Fernández Prieto<sup>3</sup>, for greater details on the Normal Bayesian partially supervised cascade classifier and on alternative initialization conditions of the iterative estimation algorithm.

## 5. A Multiple-Classifer System for a Partially Supervised Updating of Land-Cover Maps

As already pointed out, given the intrinsic complexity of the problem addressed, the partially supervised classifiers described in Sections 3 and 4 are less reliable and less accurate than the corresponding supervised classifiers. In order to alleviate this drawback, in this section, we propose an advanced classification system aimed at obtaining an accurate and robust partially supervised updating of land-cover maps. Such a system extends the approaches proposed in Sections 3 and 4, defining an effective classification framework based on a multiple classifier system (MCS). Both the above-described approaches (i.e., the independent-classification approach and the cascade-classification approach) can be considered in defining such an MCS. In the following, we give a detailed description of the system within the framework of the cascade-classification approach, thus defining a multiple cascade-classifier system (MCCS). We refer the reader to Bruzzone *et al.*<sup>6</sup> for a description of the system in the framework of partially supervised independent classification approach.

The use of an MCCS requires the definition of an ensemble of partially supervised classification techniques. In our case, the ensemble of classifiers is derived from Normal Bayesian and radial basis function (RBF) neural-network cascade-classification approaches. Three important methodological aspects are associated with the presented MCCS: i) all the partially supervised classifiers of the ensemble are defined in the framework of cascade classification; ii) a novel non-parametric partially supervised cascade classifier based on RBF neural networks is proposed; iii) hybrid Normal Bayesian and RBF neural classifiers are defined by exploiting the characteristics of the cascade-classification approach in order to generate an effective ensemble of classifiers. It is worth noting that, thanks to the non-parametric nature of the RBF neural-network cascade classifiers, the proposed system is able to analyze multisensor and/or multisource data.

This section is organized into 4 subsections. Subsection 5.1 describes the general architecture of the proposed MCCS system. Subsection 5.2 presents the partially

supervised classification problem in the framework of the cascade-classification approach for an RBF neural-network classification technique. Subsection 5.3 addresses the problem of defining suitable ensembles of cascade classifiers, and describes the proposed hybrid classifiers. Subsection 5.4 deals with the unsupervised strategies used for the combination of the results yielded by the cascade classifiers included in the considered ensemble.

### 5.1. System Architecture

The proposed system is based on a multiple cascade classifier architecture composed of  $N$  different classification algorithms (see Fig.1). The choice of this kind of architecture is due to the complexity of the partially supervised classification problem addressed. By taking into account that, in general, ensembles of classifiers are more accurate and more robust than the individual classifiers that make them up<sup>22</sup>, we expect that a multiple-classifier approach may increase the reliability and the accuracy of the global classification system.

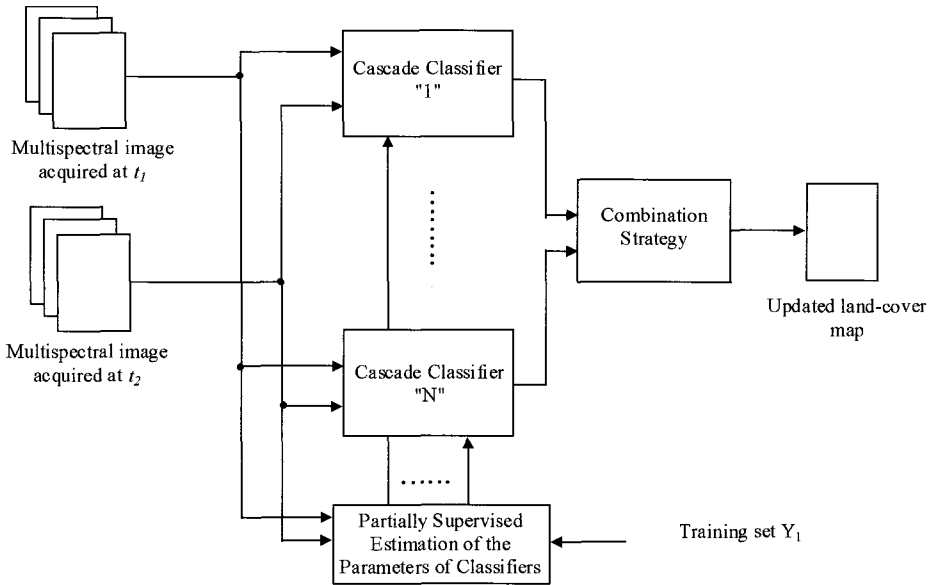


Fig. 1. General architecture of the proposed multiple-cascade-classifier system.

### 5.2. Partially Supervised Classifiers Composing the Ensemble

Let us focus our attention on the choice of each partially supervised cascade classifier to be included in the multiple-cascade-classifier architecture.

The possibility of establishing a relationship between the classifier parameters and the statistical terms involved in (12) is a basic constraint that each classification technique should satisfy in order to permit the use of the cascade-classification decision rule. To meet this requirement, we propose to use two suitable classification methods. The first is the parametric approach based on the Normal Bayesian classifier described in Section 4; the second consists of a non-parametric technique based on radial basis function (RBF) neural networks<sup>16,17</sup>. For the Normal Bayesian cascade classifier we refer the reader to subsection 4.1. Concerning the architecture of the RBF cascade classifier and the procedure for the partially supervised estimation of the related parameters, they are described in the following.

### 5.2.1. RBF Neural Network Cascade Classifier

The problem of partially supervised cascade classification by using RBF neural networks is much more complex than the one associated with the Normal Bayesian cascade classifier. The increased complexity mainly depends on the non-parametric nature of RBF neural networks. In our case, we have to resolve two critical issues in order to develop the cascade classifier in the framework of RBF neural networks: i) we should define a specific architecture that is able to implement the cascade-classification decision rule; ii) we should devise a partially supervised procedure for the training of the proposed architecture.

First of all, let us briefly recall the standard architecture of an RBF neural classifier to be used for the classification of a generic image  $\mathbf{X}_i$  (see Fig. 2). This architecture is made up of three layers: an input layer (composed of as many units as input features), a hidden layer (composed of  $S$  neurons) and an output layer (composed of as many units as land-cover classes). The input layer just propagates the input features to the hidden layer. Each unit of the hidden layer applies a simple non-linear transformation to the input data according to a symmetric radial basis function  $\varphi_s$  (usually a Gaussian function characterized by a mean value  $\pi_s$  and a width  $\sigma_s$ ). The connections between the hidden and output units are associated with a numerical value called weight (let  $w_s^r$  denote the weight that connects the  $s$ -th hidden neuron to the  $r$ -th output neuron). The output neurons apply a linear transformation to the weighted outputs of the hidden neurons. It can be proven that, if the classifier has been properly trained<sup>19</sup>, the outputs of an RBF neural network can be related to the conditional densities of the classes, which are expressed as a mixture of the kernel functions associated with the units of the hidden layer. In addition, the statistical terms computed by the neural classifier can be related to the global density function  $p(X_i)$  of the image  $\mathbf{X}_i$  as follows:

$$p(X_i) = \sum_{r=1}^C \sum_{s=1}^S p(X_i|\varphi_s)P(\varphi_s)P(\omega_r|\varphi_s) \quad (20)$$

where  $p(X_i|\varphi_s)$  is the conditional density of the variable  $X_i$  given the kernel function

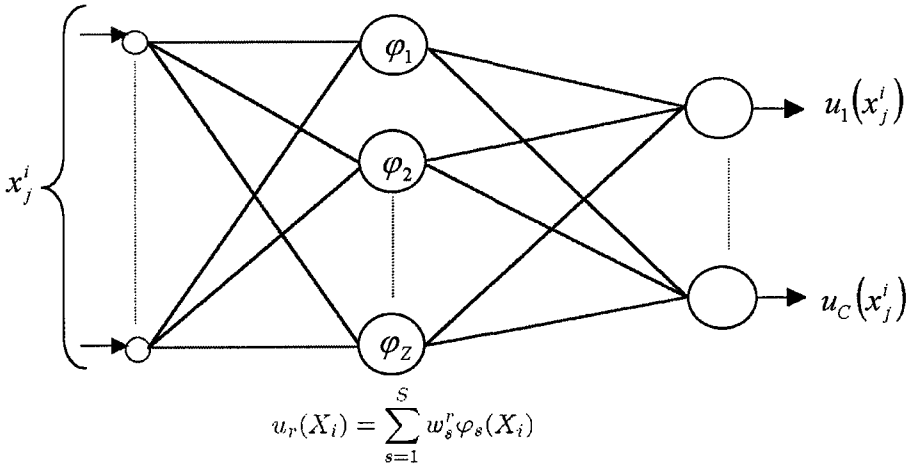


Fig. 2. Standard architecture of a supervised RBF neural-network classifier.

$\varphi_s$ ,  $P(\omega_r|\varphi_s)$  is the conditional probability of the class  $\omega_r$ , given the kernel  $\varphi_s$ ,  $P(\varphi_s)$  is the prior probability of the kernel  $\varphi_s$ , and  $S$  is the number of kernels considered. It is worth noting that the statistical terms in (20) can be associated with the parameters of the RBF neural architecture as follows<sup>19</sup>:

$$\varphi_s(X_i) = p(X_i|\varphi_s) \tag{21}$$

$$w_s^r = P(\varphi_s)P(\omega_r|\varphi_s) \tag{22}$$

We refer the reader to Bruzzone and Fernández Prieto<sup>16</sup> and to Bishop<sup>17</sup> for more details on standard RBF neural classifiers.

In order to define a cascade classifier in the context of the RBF neural-network theory, let us approximate the joint density function  $p(X_1, X_2)$  of the two images  $\mathbf{X}_1$  and  $\mathbf{X}_2$  as a mixture of Gaussian kernel functions. To this end, let us consider  $K$  kernel functions  $\varphi_k^1$  and  $Q$  kernel functions  $\varphi_q^2$  associated with the statistics of the images  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , respectively. Accordingly, under the assumption of kernel-conditional independence in the temporal domain, we can write:

$$p(X_1, X_2) \cong \sum_{h=1}^C \sum_{n=1}^C \sum_{k=1}^K \sum_{q=1}^Q p(X_1|\varphi_k^1)p(X_2|\varphi_q^2)P(\varphi_k^1, \varphi_q^2)P(\omega_n, \omega_h|\varphi_k^1, \varphi_q^2) \tag{23}$$

where  $p(X_i|\varphi_r^i)$  is the value of the conditional density function of the variable  $X_i$ , given the kernel  $\varphi_r^i$ ,  $P(\omega_n, \omega_h|\varphi_k^1, \varphi_q^2)$  is the joint conditional probability of the pair of classes  $(\omega_n, \omega_h)$  given the pair of kernels  $(\varphi_k^1, \varphi_q^2)$ , and  $P(\varphi_k^1, \varphi_q^2)$  is the joint prior

probability of the kernels  $(\varphi_k^1, \varphi_q^2)$ . In this context, the cascade classification decision rule can be rewritten as:

$$l_j^2 = \omega_m \in \Omega \text{ if and only if}$$

$$\sum_{n=1}^C \sum_{k=1}^K \sum_{q=1}^Q p(x_j^1 | \varphi_k^1) p(x_j^2 | \varphi_q^2) P(\varphi_k^1, \varphi_q^2) P(\omega_n, \omega_m | \varphi_k^1, \varphi_q^2) =$$

$$\max_{\omega_h \in \Omega} \left\{ \sum_{n=1}^C \sum_{k=1}^K \sum_{q=1}^Q p(x_j^1 | \varphi_k^1) p(x_j^2 | \varphi_q^2) P(\varphi_k^1, \varphi_q^2) P(\omega_n, \omega_h | \varphi_k^1, \varphi_q^2) \right\}. \quad (24)$$

It is worth noting that the temporal correlation between the two images is taken into account by the terms  $P(\varphi_k^1, \varphi_q^2)$  and  $P(\omega_n, \omega_h | \varphi_k^1, \varphi_q^2)$ . By analyzing equation (24), we can observe that  $p(x_j^1 | \varphi_k^1)$  and  $p(x_j^2 | \varphi_q^2)$  can be derived by applying two standard RBF neural-network classifiers to the  $t_1$  and  $t_2$  images, respectively. In particular, we can apply an RBF neural-network classifier with  $K$  hidden units to the image  $\mathbf{X}_1$  and an RBF neural-network classifier with  $Q$  hidden units to the image  $\mathbf{X}_2$  (see Fig. 3). If a proper training algorithm is used, the terms  $p(x_j^1 | \varphi_k^1)$  and  $p(x_j^2 | \varphi_q^2)$  are given by the outputs of the hidden neurons of the aforementioned neural classifiers. However, in order to implement the cascade classification decision rule, a non-conventional architecture should be considered, which involves the joint statistical terms  $P(\varphi_k^1, \varphi_q^2)$  and  $P(\omega_n, \omega_h | \varphi_k^1, \varphi_q^2)$  in the classification process. To this end, the outputs of the hidden neurons of the  $t_1$  and  $t_2$  networks are given as input to a specific block (let us call it ‘‘cascade classification’’ block) that presents as many outputs as land-cover classes (i.e.,  $C$  outputs). In particular, the output  $u_h$ , which is associated with the land-cover class  $\omega_h$ , is given by:

$$u_h(x_j^1, x_j^2) = \sum_{n=1}^C \sum_{k=1}^K \sum_{q=1}^Q P(\varphi_k^1, \varphi_q^2) P(\omega_n, \omega_h | \varphi_k^1, \varphi_q^2, \varphi_k^1(x_j^1), \varphi_q^2(x_j^2)). \quad (25)$$

According to equation (24), each pixel is classified as belonging to the land-cover class associated with the maximum output value.

The main problem that remains to be solved is the estimation of all the parameters considered in the proposed architecture in a partially supervised way (i.e., by using only the joint density function  $p(X_1, X_2)$  and the training set  $Y_1$ ). Concerning the parameters of the  $p(X_1 | \varphi_k^1)$  (i.e., the centers  $\pi_k^1$  and the widths  $\sigma_k^1$  of the Gaussian kernel functions that process the image  $\mathbf{X}_1$ ), they can be estimated according to the statistical procedure described in Bruzzone and Fernández Prieto<sup>16</sup> and in Bishop<sup>17</sup>. Consequently, the parameter vector  $\vartheta$  that remains to be estimated in a partially supervised way is composed of the following terms:

$$\vartheta = [\pi_1^2, \sigma_1^2, \dots, \pi_Q^2, \sigma_Q^2, P(\varphi_1^1, \varphi_1^2), \dots, P(\varphi_K^1, \varphi_Q^2), P(\omega_1, \omega_1 | \varphi_1^1, \varphi_1^2), \dots, P(\omega_C, \omega_C | \varphi_K^1, \varphi_Q^2)] \quad (26)$$



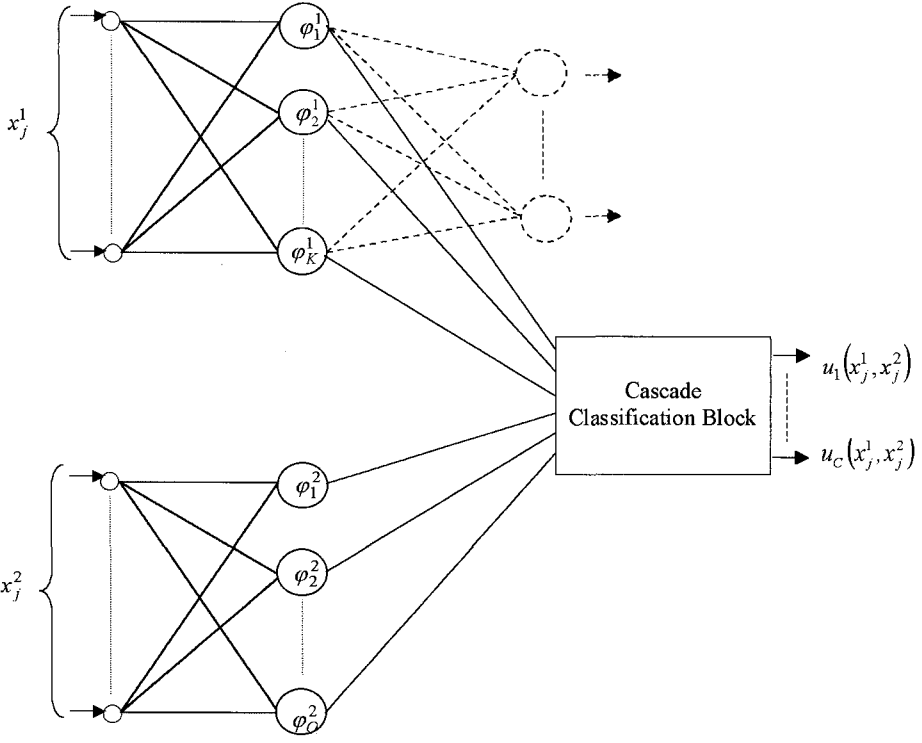


Fig. 3. Architecture of the proposed partially supervised RBF neural cascade classifier (solide line). The architecture of the standard RBF neural network used for the supervised estimation of the  $t_1$  statistical parameters is also shown (dashed line).

where  $\pi_q^2$  and  $\sigma_q^2$  are the centers and the widths characterizing the kernel functions  $\varphi_q$  that process the image  $\mathbf{X}_2$ . In order to estimate the components of the parameter vector, we propose to apply the EM algorithm to (23). Accordingly, it is possible to prove that part of the components of the parameter vector can be estimated by using the following iterative equations:

$$[\pi_q^2]^{t+1} = \frac{\sum_{j=1}^B \sum_{k=1}^K P^t(\varphi_k^1, \varphi_q^2 | x_j^1, x_j^2) x_j^2}{\sum_{j=1}^B \sum_{k=1}^K P^t(\varphi_k^1, \varphi_q^2 | x_j^1, x_j^2)} \tag{27}$$

$$[\sigma_q^2]^{t+1} = \frac{\sum_{j=1}^B \sum_{k=1}^K P^t(\varphi_k^1, \varphi_q^2 | x_j^1, x_j^2) \|x_j^2 - [\pi_q^2]^{t+1}\|^2}{d \cdot \sum_{j=1}^B \sum_{k=1}^K P^t(\varphi_k^1, \varphi_q^2 | x_j^1, x_j^2)} \tag{28}$$

$$P^{t+1}(\varphi_k^1, \varphi_q^2) = \frac{1}{B} \sum_{j=1}^B P^t(\varphi_k^1, \varphi_q^2 | x_j^1, x_j^2) \quad (29)$$

where  $d$  is the dimensionality of the input space, the superscripts  $t$  and  $t + 1$  refer to the values of the parameters at the current and next iterations, respectively, and the  $P^t(\varphi_k^1, \varphi_q^2 | x_j^1, x_j^2)$  are approximated by:

$$P^t(\varphi_k^1, \varphi_q^2 | x_j^1, x_j^2) \cong \frac{p(x_j^1 | \varphi_k^1) p^t(x_j^2 | \varphi_q^2) P^t(\varphi_k^1, \varphi_q^2)}{\sum_{z=1}^K \sum_{v=1}^Q p(x_j^1 | \varphi_z^1) p^t(x_j^2 | \varphi_v^2) P^t(\varphi_z^1, \varphi_v^2)}. \quad (30)$$

Concerning the initialization of the aforementioned components of the parameter vector  $\vartheta$ , the initial values of the parameters of the conditional density functions of kernels at  $t_2$  can be obtained by applying a standard unsupervised clustering algorithm to the  $\mathbf{X}_2$  image<sup>16</sup>, whereas the initial values of prior joint probabilities of the kernels can be easily computed in the assumption of independence between the kernels at two dates (i.e.,  $P(\varphi_k, \varphi_q) = P(\varphi_k) \cdot P(\varphi_q)$ ).

As we have already pointed out, the estimation of RBF cascade neural-network classifier parameters is significantly more complex than the estimation of Normal Bayesian cascade-classifier parameters. Despite the parameters  $\pi_q^2$ ,  $\sigma_q^2$  and  $P(\varphi_k^1, \varphi_q^2)$  of the vector  $\vartheta$  can be estimated in a fully unsupervised way, the estimation of the joint conditional probabilities  $P(\omega_n, \omega_h | \varphi_k^1, \varphi_q^2)$  requires other information in addition to the one contained in the training set  $\mathbf{Y}_1$  (it is worth noting that the terms  $P(\omega_n, \omega_h | \varphi_k^1, \varphi_q^2)$  express the relationship between kernel functions and land-cover classes). To solve this problem, we propose to exploit some of the information obtained (at convergence) by the Normal Bayesian cascade classifier described in the previous subsection. In particular, a set  $\hat{\mathbf{Y}}_2$  of pixels, which is composed of the patterns that are most likely correctly categorized by the Normal Bayesian cascade classifier, is used for the initialization of the  $P(\omega_n, \omega_h | \varphi_k^1, \varphi_q^2)$  conditional probabilities. These patterns are selected on the basis of the values of the posterior probabilities provided by the Normal Bayesian classifier. In greater detail, pixels associated with values of the posterior probabilities above a predefined threshold  $\varepsilon$  are chosen. Let  $\mathbf{Y}_{n,m}$  be the set of pairs of pixels  $(x_j^1, x_j^2)$  such that  $x_j^1 \in \mathbf{Y}_1$  belongs to the land-cover class  $\omega_n$  and  $x_j^2 \in \hat{\mathbf{Y}}_2$  is categorized by the Normal Bayesian cascade-classifier as belonging to the class  $\omega_m$ . Let  $\mathbf{Y}_{n,0}$  be the set of pairs of pixels  $(x_j^1, x_j^2)$  such that  $x_j^1 \in \mathbf{Y}_1$  belongs to the land-cover class  $\omega_n$  and  $x_j^2 \notin \hat{\mathbf{Y}}_2$ . Analogously, let  $\mathbf{Y}_{0,m}$  be the set of pairs of pixels  $(x_j^1, x_j^2)$  such that  $x_j^1 \notin \mathbf{Y}_1$  and  $x_j^2 \in \hat{\mathbf{Y}}_2$  is categorized by the Normal Bayesian cascade-classifier as belonging to the class  $\omega_m$ . The iterative equations to be used to estimate the joint conditional probabilities  $P(\omega_n, \omega_h | \varphi_k^1, \varphi_q^2)$  are the following:

$$\begin{aligned}
P^{t+1}(\omega_n, \omega_h | \varphi_k^1, \varphi_q^2) &= \frac{1}{A} \left\{ \sum_{(x_j^1, x_j^2) \in Y_{n,h}} P^t(\varphi_k^1, \varphi_q^2 | x_j^1, x_j^2) \right. \\
&+ \sum_{(x_j^1, x_j^2) \in Y_{n,0}} \sum_{g=1}^C P^t(\omega_g, \omega_h | \varphi_k^1, \varphi_q^2) P^t(\varphi_k^1, \varphi_q^2 | x_j^1, x_j^2) + \\
&\left. + \sum_{(x_j^1, x_j^2) \in Y_{0,h}} \sum_{f=1}^C P^t(\omega_n, \omega_f | \varphi_k^1, \varphi_q^2) P^t(\varphi_k^1, \varphi_q^2 | x_j^1, x_j^2) \right\}
\end{aligned} \tag{31}$$

where the normalizing factor  $A$  is equal to:

$$\begin{aligned}
A &= \sum_{h=1}^C \sum_{(x_j^1, x_j^2) \in Y_{0,h}} P^t(\varphi_k^1, \varphi_q^2 | x_j^1, x_j^2) + \sum_{n=1}^C \sum_{(x_j^1, x_j^2) \in Y_{n,0}} P^t(\varphi_k^1, \varphi_q^2 | x_j^1, x_j^2) + \\
&+ \sum_{n=1}^C \sum_{h=1}^C \sum_{(x_j^1, x_j^2) \in Y_{n,h}} P^t(\varphi_k^1, \varphi_q^2 | x_j^1, x_j^2).
\end{aligned} \tag{32}$$

It is worth noting that this iterative procedure significantly improves the initial estimates biased by the patterns included in  $\hat{\mathbf{Y}}_2$ .

Analogously to the Normal Bayesian cascade classifier, also in this case the estimated parameters evolve from their initial values to the final ones by maximizing the following log-likelihood function (the convergence to a local maximum can be proven):

$$\begin{aligned}
\ell(X_1, X_2 | \vartheta) &= \sum_{(x_j^1, x_j^2) \in \mathbf{Y}_{0,0}} \log \sum_{k=1}^K \sum_{q=1}^Q p(x_j^1 | \varphi_k^1) p(x_j^2 | \varphi_q^2) P(\varphi_k^1, \varphi_q^2) + \\
&+ \sum_{n=1}^C \sum_{(x_j^1, x_j^2) \in \mathbf{Y}_{n,0}} \log \sum_{h=1}^C \sum_{k=1}^K \sum_{q=1}^Q p(x_j^1 | \varphi_k^1) p(x_j^2 | \varphi_q^2) P(\varphi_k^1, \varphi_q^2) P(\omega_n, \omega_h | \varphi_k^1, \varphi_q^2) + \\
&+ \sum_{h=1}^C \sum_{(x_j^1, x_j^2) \in \mathbf{Y}_{0,h}} \log \sum_{n=1}^C \sum_{k=1}^K \sum_{q=1}^Q p(x_j^1 | \varphi_k^1) p(x_j^2 | \varphi_q^2) P(\varphi_k^1, \varphi_q^2) P(\omega_n, \omega_h | \varphi_k^1, \varphi_q^2) + \\
&+ \sum_{n=1}^C \sum_{h=1}^C \sum_{(x_j^1, x_j^2) \in \mathbf{Y}_{n,h}} \log \sum_{k=1}^K \sum_{q=1}^Q p(x_j^1 | \varphi_k^1) p(x_j^2 | \varphi_q^2) P(\varphi_k^1, \varphi_q^2) P(\omega_n, \omega_h | \varphi_k^1, \varphi_q^2)
\end{aligned} \tag{33}$$

where  $\mathbf{Y}_{0,0}$  is the set of pairs of pixels  $(x_j^1, x_j^2)$  such that  $x_j^1 \notin \mathbf{Y}_1$  and  $x_j^2 \notin \hat{\mathbf{Y}}_2$ .

The estimates of the parameters obtained at convergence and the ones achieved by the classical supervised procedure are used to accomplish the RBF cascade-classification process.

### 5.3. A Strategy for Generating Ensembles of Partially Supervised Cascade Classifiers: Hybrid Normal Bayesian and RBF Neural-Network Classifiers

The selection of the pool of classifiers to be integrated into the multiple cascade-classifier architecture is an important and critical task. In the literature, several

different strategies for defining a classifier ensemble have been proposed<sup>15,21,22,23,24</sup>. From a theoretical viewpoint, necessary and sufficient conditions for an ensemble of classifiers to be more accurate than any of its individual members are that the classifiers should be accurate and different<sup>25</sup>. In our case, we can control only the second condition, since no training set is available to verify the first one.

The main issue to be resolved for the definition of the ensemble concerns the capability of different classifiers to incur uncorrelated errors. In practice, several strategies have been proposed to make up pools of classifiers that incur uncorrelated errors. These strategies involve the selection of different classification algorithms, the choice of different initial training conditions for a given classification algorithm, the use of different architectures for the same kind of classifier (e.g., neural networks), the manipulation of the training examples, the manipulation of the input features, the manipulation of the output targets, the injection of randomness, etc.<sup>25</sup>. In our system, the choice of both a parametric (Normal Bayesian) and a non-parametric (RBF) classifier guarantees the use of two classification algorithms based on significantly different principles. For this reason, we expect these classifiers to incur sufficiently uncorrelated errors. However, two classification algorithms are not enough to define an effective multiple classifier architecture. To increase the reliability of the system, we need to generate a pool of  $N$  classifiers ( $N > 2$ ). According to the literature, we could define different RBF neural-network architectures in order to derive different classification algorithms for the ensemble<sup>26</sup>. However, as we are dealing with cascade-classifier techniques, we propose to adopt an alternative, deterministic, and simple strategy for making up the ensemble. This strategy is based on the characteristics of the cascade-classification approach, in which a set of key parameters, estimated by the partially supervised process, is composed of the prior joint probabilities of classes  $P(\omega_n, \omega_h)$  (they are associated with the temporal correlation between classes). The different cascade classifiers (i.e., Normal Bayesian and RBF neural networks) perform different estimations of the aforementioned probabilities, on the basis of the different classification and estimation principles. According to this observation, we propose to introduce in the ensemble hybrid classifiers obtained by exchanging the estimates of the prior joint probabilities of classes performed by different algorithms. In our case, given an Normal Bayesian cascade classifier and an RBF neural-network cascade classifier, this strategy results in an ensemble composed of the two “original” classifiers and of two hybrid Normal Bayesian and RBF algorithms obtained by exchanging the prior joint probabilities estimated in a partially supervised way by the original classifiers. These hybrid classifiers are described in the following.

Let  $P^{NB}(\omega_n, \omega_h)$ ,  $p^{NB}(X_1|\omega_n)$  and  $p^{NB}(X_2|\omega_h)$  denote the joint probabilities and the conditional densities of classes estimated by the Normal Bayesian cascade classifier, respectively. Analogously, let  $P^{RBF}(\omega_n, \omega_h|\varphi_k^1, \varphi_q^2)$ ,  $P^{RBF}(\varphi_k^1, \varphi_q^2)$ ,  $p^{RBF}(X_1|\omega_n)$  and  $p^{RBF}(X_2|\omega_h)$  denote the joint probabilities of the classes conditioned to the kernels, the joint probabilities of the kernels, and the conditional

densities of the classes at the times  $t_1$  and  $t_2$  estimated by the RBF cascade classifier, respectively.

The first hybrid classifier (let us call it Normal Bayesian-hybrid cascade classifier) is obtained by merging the joint probabilities estimated by the RBF cascade classifier with the conditional densities estimated by the Normal Bayesian cascade classifier. Hence, the corresponding classification rule is the following:

$$\begin{aligned}
 l_j^2 = \omega_m \in \Omega \text{ if and only if} \\
 \sum_{n=1}^C p^{NB}(x_j^1|\omega_n)p^{NB}(x_j^2|\omega_m)P^{RBF}(\omega_n, \omega_m) = \\
 = \max_{\omega_h \in \Omega} \left\{ \sum_{n=1}^C p^{NB}(x_j^1|\omega_n)p^{NB}(x_j^2|\omega_h) P^{RBF}(\omega_n, \omega_h) \right\}
 \end{aligned} \tag{34}$$

where:

$$P^{RBF}(\omega_n, \omega_h) = \sum_{k=1}^K \sum_{q=1}^Q P^{RBF}(\omega_n, \omega_h|\varphi_k^1, \varphi_q^2)P^{RBF}(\varphi_k^1, \varphi_q^2). \tag{35}$$

Analogously, the second hybrid classifier (let us call it RBF-hybrid cascade classifier) is obtained by merging the joint probabilities estimated by the Normal Bayesian cascade classifier with the conditional densities  $p^{RBF}(x_j^1, x_j^2|\omega_n, \omega_h)$  that can be estimated by using the RBF cascade classifier parameters. Hence, the corresponding classification rule is the following:

$$\begin{aligned}
 l_j^2 = \omega_m \in \Omega \text{ if and only if} \\
 \sum_{n=1}^C p^{RBF}(x_j^1, x_j^2|\omega_n, \omega_m)P^{NB}(\omega_n, \omega_m) = \\
 = \max_{\omega_h \in \Omega} \left\{ \sum_{n=1}^C p^{RBF}(x_j^1, x_j^2|\omega_n, \omega_h)P^{NB}(\omega_n, \omega_h) \right\}
 \end{aligned} \tag{36}$$

where the conditional densities  $p^{RBF}(x_j^1, x_j^2|\omega_n, \omega_h)$  can be approximated by:

$$\begin{aligned}
 p^{RBF}(x_j^1, x_j^2|\omega_n, \omega_h) \cong \\
 \cong \frac{\sum_{q=1}^Q \sum_{k=1}^K P^{RBF}(\omega_n, \omega_h|\varphi_k^1, \varphi_q^2)P^{RBF}(\varphi_k^1, \varphi_q^2)p^{RBF}(x_j^1|\varphi_k^1)p^{RBF}(x_j^2|\varphi_q^2)}{\sum_{q=1}^Q \sum_{k=1}^K P^{RBF}(\omega_n, \omega_h|\varphi_k^1, \varphi_q^2)P^{RBF}(\varphi_k^1, \varphi_q^2)}.
 \end{aligned} \tag{37}$$

The use of these hybrid classifiers allows one to obtain a multiple classifier architecture composed of four classifiers. It is worth noting that it is possible to further increase the number of classifiers by extending the aforementioned procedure to the case of more RBF neural network architectures with different numbers of hidden units.

#### 5.4. Multiple Cascade Classifier Architecture: Unsupervised Combination Strategies

In the proposed system, the classification results provided by the  $N$  members of the considered pool of cascade classifiers are combined by using classical multiple-classifier strategies. In particular, we consider two simple and widely used combination procedures: *Majority Voting* and *Combination by Bayesian Average*<sup>15</sup>. Both procedures exhibit the common characteristic of requiring no prior training to carry out the combination process. This is a mandatory requirement in our approach, as we have no ground truth information (and hence no training set) for the image  $\mathbf{X}_2$ . The *Majority Voting* procedure faces the combination problem by considering the results of each single classifier in terms of the class labels assigned to the patterns. A given input pattern receives  $N$  classification labels from the MCCS: each label corresponds to one of the  $C$  classes considered. The combination method is based on the interpretation of the classification label resulting from each classifier as a “vote” to one of the  $C$  land-cover classes. The data class that receives the largest number of votes is taken as the class of the input pattern.

The *Combination by Bayesian Average* strategy is based on the remark that, given the observations  $x_j^1$  and  $x_j^2$ , the  $N$  classifiers considered provide an estimate of the posterior probability  $P(\omega_h|x_j^1, x_j^2)$  for each class  $\omega_h \in \Omega$ . Therefore, a possible strategy for combining these classifiers consists in the computation of the average posterior probabilities, i.e.,

$$P^{ave}(\omega_h|x_j^1, x_j^2) = \frac{1}{N} \sum_{\gamma=1}^N \hat{P}_{\gamma}(\omega_h|x_j^1, x_j^2) \quad (38)$$

where  $\hat{P}_{\gamma}(\omega_h|x_j^1, x_j^2)$  is the estimate of the posterior probability  $P(\omega_h|x_j^1, x_j^2)$  provided by the  $\gamma$ -th classifier. The classification process is then carried out according to the Bayes rule by selecting the land-cover class associated with the maximum average probability.

## 6. Experimental Results

To assess the effectiveness of the proposed approaches, different experiments were carried out on a data set made up of two multispectral images acquired by the Thematic Mapper (TM) sensor of the Landsat 5 satellite. The selected test site was a section (412×382 pixels) of a scene including Lake Mulargias on the Island of Sardinia, Italy. The two images used in the experiments were acquired in September 1995 ( $t_1$ ) and July 1996 ( $t_2$ ). Figure 4 shows channels 5 of both images. Five land-cover classes (i.e., urban area, forest, pasture, water body, and vineyard), which characterize the test site at the above-mentioned dates, were considered. The available ground truth was used to derive a training set and a test set for each image (see Table 1). To carry out the experiments, we assumed that only the training set

associated with the image acquired in September 1995 was available. We used the training set of the July 1996 image only for comparisons with completely supervised classifiers.

Experimental results are organized into two different subsections. Subsection 6.1 presents the results obtained by the independent-classification approach proposed in Section 3. Subsection 6.2 gives the results yielded by the cascade-classification approach and the MCCS proposed in Sections 4 and 5, respectively.

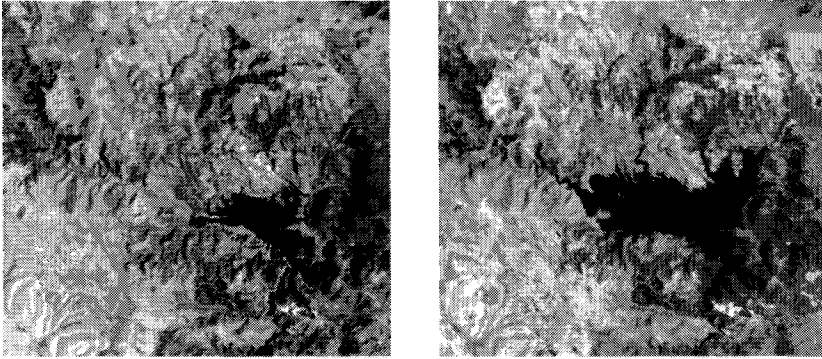


Fig. 4. Bands 5 of the Landsat-5 TM images utilized for the experiments: image acquired in September 1995 (left); image acquired in July 1996 (right).

Table 1. Number of patterns in the training and test sets for both the September 1995 and July 1996 images.

Land-cover class	Number of patterns	
	Training set	Test set
Pasture	554	589
Forest	304	274
Urban area	408	418
Water body	804	551
Vineyard	179	117
Overall	2249	1949

### 6.1. Results Obtained by the Independent-Classification Approach

The Normal Bayesian classifier was trained (in a supervised way) on the September 1995 image to estimate the *a priori* probabilities and the parameters that characterize the density functions of the classes at the time  $t_1$ . After training, the effectiveness of the classifier was evaluated on the test sets for both images. The classification

accuracies obtained are given in Table 2. On the one hand, as expected, the classifier provided a high classification accuracy (90.97%) for the test set related to the September 1995 image. On the other hand, it exhibited very poor performances for the July 1996 test set. In particular, the overall classification accuracy for the July test set was equal to 50.43%, which cannot be considered an acceptable result.

Table 2. Classification accuracies obtained for the September 1995 and July 1996 test sets by a classical supervised Normal Bayesian classifier trained on the September 1995 training set.

Land-cover class	Classification accuracy (%)	
	September 1995 test set	July 1996 test set
Pasture	82.51	19.52
Forest	97.44	95.62
Urban area	94.73	90.43
Water body	100.00	36.11
Vineyard	62.39	24.78
Overall	90.97	50.43

At this point, the proposed technique was applied to the  $t_2$  image (July 1996) in order to compute, in an unsupervised way, the new estimates of the *a priori* probabilities and density-function parameters of the considered land-cover classes. The parameters of the Normal Bayesian classifier trained on the  $t_1$  image (September 1995) were exploited to initialize the EM algorithm. At the end of the iterative process, the resulting estimates were associated with the new parameters of the Normal Bayesian classifier. In order to evaluate the accuracies of the new estimates, the classifier was tested again on the July 1996 test set. For the sake of comparison, a supervised Normal Bayesian classifier was trained and subsequently tested on the July 1996 image by using the classical approach (i.e., exploiting the training set for a supervised parameter estimation). The results obtained are given in Table 3. As one can see, the classification accuracy provided by the proposed classifier for the July test set increased by about 42%, as compared with the one exhibited by the classifier trained on the September image (92.76% vs. 50.43%). It is worth noting that this improvement was shared by most of the considered classes. A comparison with the supervised Normal Bayesian classifier trained and tested on the July image showed that such a classifier provided an overall accuracy (92.66%) very similar to the one yielded by the proposed technique (92.76%).

A further insight into the behavior of the proposed method is provided by Fig. 5, where the trend of the overall classification accuracy versus the number of EM-algorithm iterations is plotted. As can be seen, the overall classification accuracy increases significantly from 50.43% (i.e., for the initial estimates) to 88.19% in only 10 iterations, and reaches the final value of 92.76% in 23 iterations.



Table 3. Classification accuracies obtained for the July 1996 test set by the proposed Normal Bayesian classifier retrained on the July 1996 image. For the sake of comparison, the classification accuracies achieved by a classical supervised Normal Bayesian classifier trained and tested on the July 1996 image are also given.

Land-cover class	Classification accuracy (%)	
	Proposed independent partially supervised technique	Supervised Normal Bayesian classifier trained on the July training set
Pasture	94.06	92.02
Forest	87.22	92.70
Urban area	93.06	93.30
Water body	100.00	100.00
Vineyard	64.10	58.97
Overall	92.76	92.66

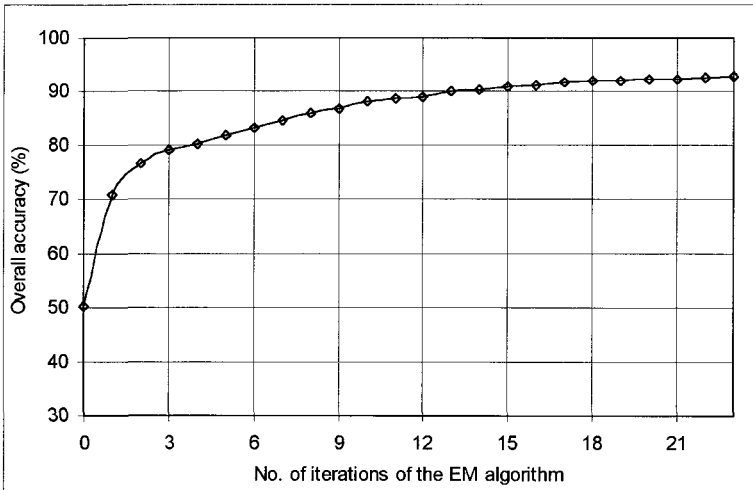


Fig. 5. Classification accuracy obtained by the proposed Normal Bayesian classifier for the July 1996 test set versus number of iterations of the EM algorithm.

**6.2. Results Obtained by the Cascade-Classification Approach and the MCCS**

Partially supervised Normal Bayesian and RBF neural-network cascade classifiers were applied to the September 1995 and July 1996 images. For the Normal Bayesian cascade classifier, the assumption of Gaussian distributions was made for the density functions of the classes (this is a reasonable assumption as we considered TM images). Concerning the RBF neural cascade classifier, in order to exploit its non-parametric nature, five texture features based on the Gray-Level Co-occurrence matrix (i.e., sum variance, sum average, correlation, entropy and difference variance)<sup>28</sup> were computed and given as input to the classifier in addition to the six TM chan-

nels. These features were obtained by using a window size equal to  $7 \times 7$  and an interpixel distance equal to 1.

As regards the Normal Bayesian cascade classifier, the parameters of the Gaussian density functions of the classes at  $t_1$  were computed in a supervised way by using the available training set for the September 1995 image (i.e.,  $\mathbf{Y}_1$ ). These values were also used to initialize the parameters of the conditional density functions of the classes at  $t_2$ . Concerning the RBF cascade classifier, several trials were carried out in order to derive an effective number of neurons to be used in the hidden layer. To this end, experiments were carried out using a standard RBF architecture trained by the available set  $\mathbf{Y}_1$  and applied to the  $t_1$  test set. The highest accuracy was obtained by an architecture composed of 35 hidden units. On the basis of this result, an architecture composed of 70 hidden units was used for the RBF cascade classifier (i.e., 35 units related to the  $t_1$  image and 35 units related to the  $t_2$  image). It is worth noting that the parameters of the 35 hidden units associated with  $\mathbf{X}_1$  were fixed according to the values achieved in a supervised way in the aforementioned experiment. The values of the parameters of the 35 hidden units used to process the image  $\mathbf{X}_2$  were initialized by applying an unsupervised clustering to that image.

The parameters of the vectors  $\vartheta$  related to the Normal Bayesian and RBF cascade classifiers were estimated in an unsupervised way by using the proposed formulations of the iterative EM algorithm (see (15)-(17), and (27)-(32)). Firstly, the Normal Bayesian cascade classifier was trained, and the patterns classified with a posterior probability higher than the threshold value  $\varepsilon=0.98$  were used to generate the set  $\hat{\mathbf{Y}}_2$  in order to support the RBF training process. The EM algorithms adopted for the Normal Bayesian and RBF partially supervised training processes converged in 11 and 25 iterations, respectively. At the end of the iterative process, the resulting estimates were used to perform the classification of the July 1996 image. In addition, from the considered Normal Bayesian and RBF cascade classifiers, the two hybrid Normal Bayesian and RBF neural-network cascade classifiers were derived according to the strategy described in subsection 5.3. Also these hybrid classifiers were applied to the July 1996 image.

The classification accuracies exhibited by the aforementioned four partially supervised cascade classifiers on the  $t_2$  test set are given in Table 4. As one can see, the performances of all the classifiers are very good. In particular, the overall accuracies exhibited by both the RBF and RBF-hybrid classifiers are very high (i.e., 96.10% and 95.38%, respectively), and also the overall accuracies provided by the Normal Bayesian and Normal Bayesian-hybrid classifiers are satisfactory (i.e., 91.48% and 91.79%, respectively). This confirms the effectiveness of the partially supervised training process. Comparisons between standard and hybrid classifiers (i.e., RBF vs. RBF-hybrid and Normal Bayesian vs. Normal Bayesian-hybrid) point out that these classifiers provided very similar overall accuracies. However, a deeper analysis of the results reveals some important differences between the considered classification techniques. For example, the accuracy exhibited by the RBF-hybrid cascade

classifier on the vineyard class is significantly higher than the one exhibited by the RBF neural cascade classifier (i.e., 66.67% vs. 61.54%). If one considers the confusion matrices resulting from the aforementioned experiments (see Tables 5 (a)-(d)), one can verify other significant differences in the behaviors of the classifiers on the different classes. For example, the RBF classifier misclassifies 30 pasture patterns as belonging to urban areas, whereas the RBF-hybrid classifier never incurs such a classification error. This confirms that the assumption that the four classifiers incur quite uncorrelated errors is reasonable.

Table 4. Classification accuracies obtained by the four partially supervised cascade classifiers included in the proposed multiple classifier architecture (July 1996 test set).

Land-cover class	Classification accuracy (%)			
	Normal Bayesian	RBF	Normal-Bayesian-hybrid	RBF-hybrid
Pasture	83.53	94.91	85.23	94.40
Forest	97.45	100.00	97.45	98.91
Urban area	95.69	99.76	94.98	96.41
Water body	100.00	100.00	100.00	100.00
Vineyard	62.39	61.54	61.54	66.67
Overall	91.48	96.10	91.79	95.38

At this point, the four classifiers were combined by using both the Majority-Voting and the Combination by Bayesian Average strategies (concerning the Majority-Voting strategy, in the case where more than one class received the same maximum number of votes, the class with the maximum posterior probability was chosen). The accuracies obtained on the July 1996 test set are given in Table 6. Both combination strategies provided very high accuracies on all the land-cover classes, with the exception of the vineyard class, which is a minority one. By comparing Tables 4 and 6, one can conclude that the classification accuracies obtained combining the results of the partially supervised cascade classifiers by the two combination strategies considered are significantly higher than the accuracy exhibited by the worst single classifier (i.e., 96.56% and 94.77% vs. 91.48%). In particular, the classification accuracy obtained by applying the majority rule strategy is also higher than those exhibited by all the single classifiers making up the ensemble.

As stated in the methodological part of the chapter, the objective of the multiple-classifier approach is not only to improve the overall classification accuracy of the system but also to increase its robustness. In order to investigate this aspect, an experiment was carried out in which the failure of the training process of the RBF neural cascade classifier was simulated. In particular, in order to simulate this situation, the partially supervised training of the parameters of the RBF architecture was carried out by replacing the image  $\mathbf{X}_2$  with the image  $\mathbf{X}_1$ . It is worth noting that the resulting incorrect estimation of the RBF parameters also affects the hybrid classifiers. Table 7 presents the classification accuracies obtained by this experiment. As can be seen, even though the overall accuracies exhibited by both the RBF and the

Table 5. Confusion matrices that resulted from the classification of the July 1996 test set by using the proposed partially supervised techniques: a) Normal-Bayesian cascade classifier; b) RBF neural cascade classifier; c) Normal-Bayesian-hybrid cascade classifier; d) RBF-hybrid neural cascade classifier.

	Pasture	Forest	Urban area	Water body	Vineyard
Pasture	492	12	85	0	0
Forest	2	267	2	0	3
Urban area	5	5	400	0	8
Water body	0	0	0	551	0
Vineyard	23	11	10	0	73

(a)

	Pasture	Forest	Urban area	Water body	Vineyard
Pasture	559	0	30	0	0
Forest	0	274	0	0	0
Urban area	0	0	417	1	0
Water body	0	0	0	551	0
Vineyard	31	11	3	0	72

(b)

	Pasture	Forest	Urban area	Water body	Vineyard
Pasture	502	15	72	0	0
Forest	2	267	2	0	3
Urban area	5	7	397	0	9
Water body	0	0	0	551	0
Vineyard	21	11	13	0	72

(c)

	Pasture	Forest	Urban area	Water body	Vineyard
Pasture	556	23	0	10	0
Forest	0	271	0	2	1
Urban area	15	0	403	0	0
Water body	0	0	0	551	0
Vineyard	21	0	3	15	78

(d)

Table 6. Overall classification accuracies exhibited by the proposed multiple cascade classifier system.

Land-cover class	Classification accuracy (%)	
	Bayesian Average	Majority rule
Pasture	91.51	94.06
Forest	99.27	99.64
Urban area	98.09	99.28
Water body	100.0	100.0
Vineyard	64.10	76.06
Overall	94.77	96.56

RBF-hybrid cascade classifiers are very poor (i.e., 67.68% and 72.75%, respectively), both combination strategies (i.e., the Combination by Bayesian Average strategy and the majority rule) allow the presented system to achieve classification accuracies (i.e., 92.46% and 95.90%) higher than the ones yielded by all the single classifiers. This confirms that the proposed architecture based on multiple cascade classifiers permits one to increase the robustness of the system versus possible failures of the partially supervised training process of single cascade-classification techniques.

Table 7. Overall classification accuracies exhibited by the four partially supervised cascade-classifiers included in the proposed multiple classifier architecture (July 1996 test set). The results are related to the case in which a failure in the partially supervised training of the RBF cascade-classifier was simulated. The overall accuracy obtained after combining the proposed classifiers is also given.

		Classification accuracy (%)			
Normal Bayesian	RBF	RBF-hybrid	Normal Bayesian-hybrid	Bayesian average	Majority rule
91.48	67.68	72.75	91.74	92.46	95.90

Finally, in order to completely assess the effectiveness of the proposed methodology, two additional experiments were carried out using a fully supervised standard RBF classifier. In the first experiment, the RBF classifier was trained on the September 1995 training set and tested on the July 1996 image. The obtained results are given in Table 8. As one can see, the standard supervised RBF neural-network classifier trained on the “old” training set was unable to classify the “new” image with an acceptable accuracy, thus confirming that the use of a more complex classification methodology based on a partially supervised training process is mandatory. In the second experiment, the RBF classifier was trained on the July 1996 training set and applied to the test set related to the same image (it is worth noting that this training set was not considered in the previous experiments as we assumed that it was not available). Table 9 gives the obtained results. A comparison of these results with the ones provided in Table 6 points out that the proposed system outperforms the standard supervised RBF classifier. This surprising result, which mainly depends on the ability of the proposed approach to exploit the temporal correlation between the two images considered, confirms the effectiveness of the presented methodology.

## 7. Discussion and Conclusions

In this chapter, partially supervised approaches to classification of multitemporal remote-sensing images have been presented. The proposed approaches allow one to produce accurate land-cover maps of a specific study area also from images for which a reliable ground truth (hence a suitable training set) is not available. They can be used in applications in which the area of interest is characterized by the same kinds of land-cover classes over time. This means that only the spatial distributions

Table 8. Classification accuracies exhibited by a standard supervised RBF neural classifier trained on the September 1995 image and tested on the July 1996 image.

Land-cover class	Classification accuracy (%)
Pasture	47.70
Forest	94.16
Urban area	66.27
Water body	100.00
Vineyard	45.30
Overall	72.85

Table 9. Classification accuracies exhibited by a standard supervised RBF classifier trained and tested on the July 1996 image.

Land-cover class	Classification accuracy (%)
Pasture	89.64
Forest	99.27
Urban area	88.28
Water body	100.00
Vineyard	67.52
Overall	92.30

of land-covers are assumed to change over time.

The presented methods are based on the assumption that the estimates of the distributions of classes derived from a supervised training on a previous image of the considered area can represent rough estimates of the class distributions in the new image to be categorized. Then the EM algorithm is applied in order to improve such estimates iteratively on the basis of the global density function of the new image. It is worth noting that when the initial estimates are very different from the true ones (e.g., when the considered image has been acquired under atmospheric or light conditions very different from the ones in the image exploited for the supervised initial training of the classifier), the EM algorithm may lead to inaccurate final values. Therefore, in order to overcome this problem, we recommend the application of a suitable pre-processing phase aimed at reducing the main differences between images due to the above-mentioned factors (simple relative-calibration techniques, which do not require any atmospheric data, can be adopted<sup>29,30</sup>). In addition, the sequence of images to be classified should be acquired in similar periods of the year, as the spectral responses of the related land covers (hence the corresponding distributions in the feature space) may significantly change in different seasons.

Experiments carried out on different multitemporal data sets confirmed the validity of the proposed techniques (in this chapter only experiments carried out on a data set have been reported). In particular, the resulting classifiers revealed very effective, and attained high classification accuracies for the new images, without

relying on the corresponding training sets.

It is worth noting that, despite in the results given in Section 6 the accuracies yielded by the independent-classification approach and by the cascade-classification approach are similar, in general it is reasonable to expect that the cascade-classification approach results more reliable than the independent-classification one, thanks to its ability to exploit the temporal correlation between the two considered images.

Although the independent and the cascade approaches have proven effective, they exhibit some limitations. In particular, given the intrinsic complexity of the problem addressed, these approaches may result in classifiers less reliable and accurate than the corresponding supervised classifiers. To overcome this limitation, the use of an ensemble of partially supervised classifiers has been proposed. In this chapter, a multiple classifier architecture has been presented in the framework of the cascade classification. The main features exhibited by the resulting system are: i) robustness to the partially supervised training process, thanks to the use of different partially supervised classifiers; ii) capability to consider multisensor and multisource data in the process of updating of land-cover maps (thanks to the availability of non-parametric classification algorithms in the ensemble).

The proposed approaches seem very promising tools to be integrated into a GIS for a regular updating of land-cover maps. It is worth noting that, in the case where an “old” ground truth is not available, the land-cover map itself at time  $t_1$  can be considered as the training set  $\mathbf{Y}_1$  required for the partially supervised training process of the proposed system (however, in this situation, the possible errors present in the original land-cover map may affect the accuracy of the system).

## Acknowledgments

This research was supported in part by the Italian Space Agency and in part by the Italian Ministry of Education, University and Research.

## References

1. J. A. Richards, *Remote sensing digital image analysis*, 2<sup>nd</sup> ed., New York: Springer-Verlag, 1993.
2. P.H. Swain, “Bayesian classification in time-varying environment,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, pp. 880-883, 1978.
3. L. Bruzzone and D. Fernández Prieto, “A partially unsupervised approach to the automatic classification of multitemporal remote-sensing images,” *Pattern Recognition Letters*, vol. 23, pp. 1063-1071, 2002.
4. L. Bruzzone and R. Cossu, “A Multiple Cascade-Classifer System for a Robust and Partially Unsupervised Updating of Land-Cover Maps,” *IEEE Trans. on Geosci. Remote Sens.*, vol. 40, no. 9, pp. 1984-1996, 2002.
5. L. Bruzzone and D. Fernández Prieto, “Unsupervised retraining of a maximum-likelihood classifier for the analysis of multitemporal remote-sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 456-460, 2001, 2001.

6. L. Bruzzone, R. Cossu, G. Vernazza, "Combining parametric and non-parametric algorithms for a partially unsupervised classification of multitemporal remote-sensing images", *Journal on Information Fusion*, in press.
7. F. Maselli, M. A. Gilabert, and C. Conese, "Integration of high and low resolution NDVI data for monitoring vegetation in mediterranean environments," *Remote Sensing of Environment*, vol. 63, pp. 208-218, 1998.
8. A. Grignetti, R. Salvatori, R. Casacchia, and F. Manes, "Mediterranean vegetation analysis by multi-temporal satellite sensor data," *Int. J. Remote Sensing*, vol. 18, no. 6, pp. 1307-1318, 1997.
9. M.A. Friedl, C. E. Brodley, and A. H. Strahler, "Maximizing land cover accuracies produced by decision trees at continental to global scales," *IEEE Trans. on Geosci. Remote Sens.*, vol. 37, no. 2, pp. 969-977, 1999.
10. J.T. Tou and R.C. Gonzalez, *Pattern recognition principles*. Reading, MA: Addison, 1974.
11. B. Jeon and D. Landgrebe, "Partially supervised classification used weighted unsupervised clustering," *IEEE Trans. on Geosci. Remote Sens.*, vol. 37, no. 2, pp. 1073-1079, 1999.
12. A. P. Dempster, N.M. Laird and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of Royal Statistic. Soc.*, vol. 39, no. 1, pp. 1-38, 1977.
13. B.M. Shahshahani and D. Landgrebe, "The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon," *IEEE Trans. on Geosci. Remote Sens.*, vol. 32, no. 5, pp. 1087-1095, 1994.
14. T.K. Moon, "The Expectation-Maximization algorithm", *Signal Processing Magazine*, vol. 13, no. 6, pp. 47-60, 1996.
15. J. Kittler, A. Hojjatoleslami, and T. Windeatt, "Strategies for combining classifiers employing shared and distinct pattern representations," *Pattern Recognition Letters*, vol. 18, pp. 1373-1377, 1997.
16. L. Bruzzone and D. Fernández Prieto, "A technique for the selection of kernel-function parameters in RBF neural networks for classification of remote-sensing images," *IEEE Trans. on Geosci. Remote Sens.*, vol. 37, no.2, pp. 1179-1184, 1999.
17. C.M. Bishop, *Neural networks for pattern recognition*. Oxford: Clarendon Press, 1995.
18. L. Bruzzone, D. Fernández Prieto, and S. B. Serpico, "A neural statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Trans. on Geosci. Remote Sens.*, vol. 37, no.3, pp.1350-1359, 1999.
19. D.J. Miller and S.U. Hasan, "Combined learning and use for a mixture model equivalent to the RBF classifier," *Neural Computation*, vol. 10, pp. 281-293, 1998.
20. L. Bruzzone and S.B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, pp. 858-867, 1997.
21. L. Lam and C.Y. Suen, "Application of majority voting to pattern recognition: An analysis of its behavior and performance," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 27, pp. 553-568, 1997.
22. J. Kittler, M. Hatef, R.P.W. Duin, and J. Mates, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 126-239, 1998.
23. J.A. Benediktsson and P.H. Swain, "Consensus theoretic classification methods," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, pp. 688-704, 1992.
24. J.A. Benediktsson, J.R. Sveinsson, O.K. Ersoy, and P.H. Swain, "Parallel consensual neural networks," *IEEE Transactions on Neural Networks*, vol. 8, pp. 54-64, 1997.



25. T.G. Dietterich, "Ensemble methods in machine learning," in *Proc. of 1st International Workshop on Multiple Classifier Systems*, Cagliari, Italy, pp. 1-15, 2000.
26. A.J.C. Sharkey, N.E. Sharkey, U. Gerecke, and G.O. Chandroth, "The test and select approach to ensemble combination," in *Proc. of 1st International Workshop on Multiple Classifier Systems*, Cagliari, Italy, pp. 30-44, 2000.
27. L. Bruzzone, F. Roli, and S.B. Serpico, "An extension of the Jeffreys-Matusita distance to multiclass cases for feature selection", *IEEE Trans. on Geosci. Remote Sens.*, vol. **33**, no. 6, pp. 1318-1321, 1995.
28. R.M. Haralick, K. Shanmugan, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. **3**, pp. 610-621, 1973.
29. P. S. Chavez, Jr., "Radiometric calibration of Landsat Thematic Mapper multispectral images," *Photogrammetric Engineering & Remote Sensing*, vol. **55**, no. 9, pp. 1285-1294, 1989.
30. P. S. Chavez, Jr. and D. J. MacKinnon, "Automatic detection of vegetation changes in the southwestern United States using remotely sensed images," *Photogramm. Eng. Remote Sens.*, vol. **60**, no. 5, pp. 1285-1294, 1994.

# CHAPTER 13

## NEAREST NEIGHBOR DECISION RULE FOR PIXEL CLASSIFICATION IN REMOTE SENSING

Szymon Grabowski<sup>1</sup>, Adam Jóźwik<sup>1,2</sup>, C.H. Chen<sup>3</sup>

<sup>1</sup>*Technical University of Łódź, Computer Engineering Department  
Al. Politechniki 11, 90-924 Łódź, E-mail: [SGrabow@zly.kis.p.lodz.pl](mailto:SGrabow@zly.kis.p.lodz.pl)*

<sup>2</sup>*Institute Biocybernetics and Biomedical Engineering, Polish Academy of Sciences  
02-109 Warsaw, Trojdena 4*

<sup>3</sup>*University of Massachusetts Dartmouth, Electrical and Computer Engineering Dept.,  
285 Old Westport Road, N.Dartmouth*

A classifier based on the  $k$ -NN rule is known as the one that offers a very good performance. Learning of such kind of classifiers consists in determination the value of  $k$ . Some modification of the standard  $k$ -NN rule may lead to the improvement of the classification quality. The relatively new  $k$  Nearest Centroid Neighbor ( $k$ -NCN) decision rule uses an interesting concept of surrounding neighborhood, that is such a neighborhood, which takes into account not only the proximity of neighbors, but also their spatial location. Neighbors should be located not only close to a query sample, but also possibly around it in the space. In this chapter we present our decision rule called  $k$  Near Surrounding Neighbors ( $k$ -NSN), which “improves” the neighborhood used in  $k$ -NCN with respect to both described aspects. Moreover, we present a voting technique which finds several  $k$  parameters for  $k$ -NN,  $k$ -NCN and  $k$ -NSN rules learnt from random partitions of the training set and utilizes them in an ensemble of classifiers. As opposed to most ensemble methods, our algorithms require moderate computational increase in relation to the base classifiers, and even almost negligible computation increase in the voting  $k$ -NN case. We test the aforementioned methods on a remote sensing dataset (already used in several experiments) and obtain results which show attractiveness of the presented concepts in applications where prediction accuracy is of primal importance. The main disadvantage of the  $k$ -NN decision rule and its modified versions is a necessity of keeping the whole training set, as the reference set, in the computer memory during a classification phase. Numerous procedures, which have been already proposed for reference set reduction, concern the 1-NN rule. Although most proposed methods were originally devised for the 1-NN rule, there is no obstruction to use the received reduced sets with  $k$ -NN classifiers. It is also possible to reclassify the original reference set by applying the  $k$ -NN rule, standard or modified, and then to use the simple 1-NN rule with the reclassified set. The effectiveness of these approaches will be studied in relation to four different algorithms of reference set size reduction.

### 1 Introduction

Remote sensing image analysis, in its final stage, consists in classification of pixels. The construction of a classifier is based on the large training set. Furthermore, not only a good performance but also the speed of classification phase plays a very important role in a choice of the classifier type. The best possible decision rule offers the Bayes classifier that operates according to the formula:  $p(j/\mathbf{x})=p(j)\cdot f(\mathbf{x}/j)/f(\mathbf{x})$ , where  $p(j/\mathbf{x})$  is a probability of the class  $j$  under assumption that the classified object is described by a feature vector  $\mathbf{x}$ ,  $f(\mathbf{x}/j)$  denotes the density of probability distribution for the class  $j$  and  $f(\mathbf{x})$  is the density of probability distribution of the feature vector  $\mathbf{x}$ . The vector  $\mathbf{x}$  is assigned to the class  $j$  that corresponds to the maximum value of  $p(j/\mathbf{x})$ . For convenience, the feature vectors  $\mathbf{x}$  will be treated also as points in the feature space.

All functions which appear on the right side in the above mentioned Bayes formula are unknown. To approximate them one can use the neighborhood containing the  $k$  nearest neighbors of the classified point  $\mathbf{x}$ . In this way,  $p(j)\approx m_j/m$ ,  $f(\mathbf{x}/j)\approx k_j/(m_j\cdot V(\mathbf{x},k))$ ,  $f(\mathbf{x})\approx k/(m\cdot V(\mathbf{x},k))$ , where  $m_j$  is a number of objects from the class  $j$  in the training set,  $m$  is a numerical force of the training set,  $k_j$  means a number of points from the class  $j$  among  $k$  nearest points (neighbors) of the classified point  $\mathbf{x}$  and  $V(\mathbf{x},k)$  is a volume occupied by a

hypersphere containing these  $k$  nearest neighbors. Thus, the left side can be calculated. The probability function that occurs on the left side of the Bayes formula is then approximated by the ratio  $k_j/k$ , i.e.  $p(j/\mathbf{x}) \approx k_j/k$ . In this way the  $k$  nearest neighbor ( $k$ -NN) decision rule has been obtained. The classified point  $\mathbf{x}$  is assigned to the class  $j$  that corresponds to the highest value of  $k_j/k$ . It is proved by others authors [1] that if the size  $m$  of the training set gets larger ( $m \rightarrow \infty$ ),  $k \rightarrow \infty$  and  $k/m \rightarrow 0$  then the performance of the  $k$ -NN rule converges to the performance of Bayes' classifier. That is why the classifier based on  $k$ -NN rule has been chosen as a subject of the present paper. The  $k$ -NN rule was originally proposed in [2].

The training set containing points with known class membership is the set that is used for the classifier construction. In the case of the  $k$ -NN rule, it may serve for experimental determination a value of the parameter  $k$ . For instance, we can use the well known *leave-one-out* misclassification rate estimation or a cross-validation technique [3] to select the value of  $k$  which offers the best performance.

In recent years the concept of so-called *surrounding neighborhood* has been introduced. Such neighborhood can intuitively be understood as an item subject to two complementary constraints. Firstly, the neighbors of a query point  $q$  should be as **close** to it as possible. Secondly, the neighbors should also be located as **symmetrically** around  $q$  as possible. The  $k$ -NN ignores the latter aspect. Let us concentrate on one practical proposal fitting into this framework.

The idea of Chaudhuri [4], later developed to a decision rule [5], seems to be an interesting attempt to focus on neighbors, which are located not only close enough to the given sample, but also possibly homogeneously distributed around the sample. It is the concept of Nearest Centroid Neighborhood (NCN). The  $k$  nearest centroid neighbors of a query point  $q$  are obtained as follows:

- first neighbor of  $q$  is its nearest neighbor,  $n_1$ ;
- the  $i$ 'th neighbor,  $n_i$ ,  $i \geq 2$ , is such that the centroid (i.e. the mean)  $c_i$  of this and previously selected neighbors,  $n_1, \dots, n_{i-1}$ , is the closest to  $q$ .

Because of the centroid criterion, the spatial distribution of neighbors is taken into account. On the other hand, the incremental nature of the way in which successive neighbors are obtained guarantees their proximity to the query sample  $q$ .

Experiments conducted by Sánchez et al. ([5, 6]<sup>1</sup>) confirm attractiveness of the  $k$ -NCN decision rule, especially in applications where classification accuracy is more important than classification time. The  $k$ -NCN usually outperformed the standard  $k$ -NN rule.

Just like for  $k$ -NN, also for  $k$ -NCN the number of neighbors must be estimated with respect to the training set, preferably with the leave-one-out method.

In [7] we proposed another surrounding neighborhood based decision rule, called  $k$  Near Surrounding Neighborhood ( $k$ -NSN), which tries to optimize both criteria used by  $k$ -NCN.

In this chapter we propose an ensemble of  $k$ -NN (or  $k$ -NCN, or  $k$ -NSN...) classifiers in which the values of  $k$  are diversified via estimations performed on various random partitions of the training set. We believe such an approach is, to a certain degree, a protection from overfitting. An experimental confirmation of the idea is presented in

---

<sup>1</sup> In our experiments on the remote sensing dataset used in the paper and on several UCI datasets, the better of the two modifications to the  $k$ -NCN rule proposed in [6], offered a notably worse accuracy than the original rule.

Section 5. Moreover, as opposed to most classification schemes with voting over multiple classifiers, the proposed technique does not require an increase in computational resources directly proportional to the number of components.

## 2 The $k$ Near Surrounding Neighbors ( $k$ -NSN) decision rule

The heuristic nature of  $k$ -NCN encourages us to search for other decision rules which would take into account both described aspects of neighborhood. As the centroid criterion from  $k$ -NCN seems really good, we decided not to change it, but only optimize the set of neighbors according to both criteria: one related to proximity and the other to the distance of the neighbor set's centroid to the test sample.

Our classifier first searches for  $k$  NCN neighbors of a test sample and then in a loop tries to exchange some neighbors with other samples which are both closer and better in the sense of the centroid criterion to the test sample.

More precisely, the proposed decision rule, which we called  $k$  Near Surrounding Neighbors ( $k$ -NSN), operates as described in Fig. 1.

Note it is a *random mutation hill climbing* algorithm; the neighbor exchanging idea is analogous to the one applied by Skalak to prototype selection [8].

In our experiments the learning phase for  $k$ -NSN was performed with use of the  $k$ -NCN rule, which is faster.

---

```

Find  $k$  nearest centroid neighbors of a test sample  $q$ . Call the neighbors  $n_1, \dots, n_k$ . Call
their centroid  $c$ .
For  $counter = 1$  to ITERATIONS do
{
  Select a random neighbor  $n_i, 1 \leq i \leq k$ 
  Select a random sample  $s$ , such that  $d(q, s) \leq d(q, n_{farthest})$ , where  $n_{farthest}$  is the  $k$ 'th,
  according to the distance, NCN neighbor of  $q$ 
  If  $d(q, s) \leq d(q, n_i)$ 
  {
    Let  $tentative\_c = \text{centroid}(n_1, \dots, n_{i-1}, s, n_{i+1}, \dots, n_k)$ 
    If  $d(q, tentative\_c) < d(q, c)$ 
    {
      Let  $n_i = s$ 
      Let  $c = tentative\_c$ 
    }
  }
}
}
Return the set of neighbors  $n_1, \dots, n_k$ .

```

---

Figure 1: The  $k$  Near Surrounding Neighbors ( $k$ -NSN) decision rule

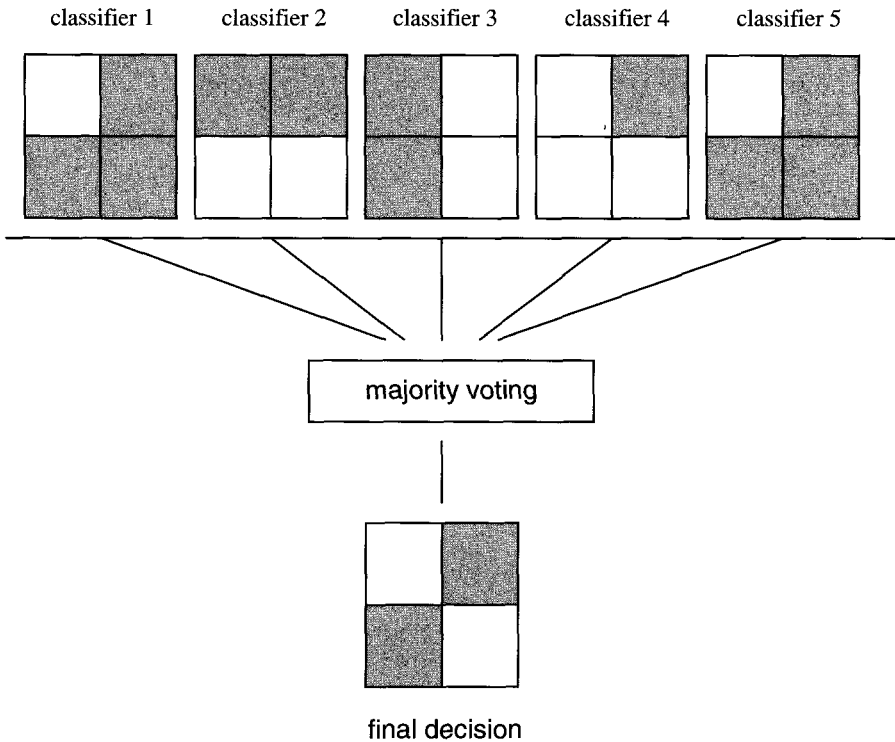
## 3 Voting over multiple classifiers

An essential problem in all classification tasks is the danger of overfitting. This phenomenon consists in choosing the classifier and its parameters (comprising e.g. some numeric values, its reference set(s) *etc.*) so well fitting the learning data that it does not actually fit the real concept. The trouble „as such” is inevitable since in practice we always deal with finite datasets, however ways to mitigate it have been developed.

One of the most attractive approaches to increase accuracy and/or minimize error variance (i.e. generate classification models which are more trustworthy to be safe from excessive overfitting) is the idea of combining classifiers. It has attained a vivid interest in the past decade, which resulted in a number of theoretical and practical achievements, especially in the domain of decision trees (for a survey see e.g. [9, 10]).

The success of ensembles can be explained in following words: if each particular voter (=component classifier) produces different approximation to real decision boundaries, then during the voting (which is kind of averaging process) the noise is supposed to be smoothed out, especially if the number of voters is large. This, however, holds true when the voters are independent. The condition of independence is not only hard to fulfill but even to measure in a real-life (finite) task. In practice we thus rather strive to select weakly correlated (“diverse”), but still quite accurate classifiers. Several heuristic measures of diversity between component classifiers have been proposed and tested [11, 12], but any stronger guidelines on how to choose the classifiers for the ensemble still cannot be given.

Fig. 2 shows a case where an ensemble of five simple classifiers under majority voting solves the classic XOR problem. Note the components are not independent (e.g. classifiers 1 and 5 have same predictions on the whole domain).



**Figure 2: An ensemble of classifiers solves the XOR problem**

Little has been done specifically in combining nearest neighbor classifiers. We are aware only of a few papers in this domain, namely Skalak’s combining 1-NN classifiers with radically reduced prototype sets [8, 13], voting over Hart’s condensed nearest neighbor classifiers [14], Multiple Feature Subsets (MFS) algorithm [15] and

decomposition of a multi-class problem into a net of dichotomizers [16, 17] (the last idea is of more general use and in fact has been known for years in neural network community [1]).

In this chapter we attempt to overcome the problem of selecting  $k$  in  $k$ -NN (or another  $k$  neighbors' based classifier) through using a number of  $k$ -NN (or such like) classifiers, each working on the whole reference set, but with its own value of  $k$ . Final decision is obtained via simple voting. The component  $k$ -NN classifiers are trained on random partitions of the whole learning set. Our goal is to increase accuracy rather than to decrease training costs (which in fact must be greater in our version compared to plain  $k$ -NN). Beneath we refer to the algorithm as to *voting  $k$ -NN*.

How does the learning in voting  $k$ -NN proceed in detail? A known technique which allows to estimate a classifier during the learning stage (i.e. in design time) is partitioning the whole learning set into two parts: a „real” training set and a validation set. The models for the classifier under design are applied for the former set but tested on the latter one. Of course, an obvious deficiency of such a technique is shrinking the training set. This, however, seems a price we must pay for the comfort of some estimation of the classifier yet in the learning stage. The question arises: how the given set should be divided into those two parts? Too few samples in the training set results in a very weak approximation of the underlying distribution. Too few samples in the validation set, on the other hand, implies an unreliable estimation of the generated classifier.

Our decision was to divide the learning set into halves. The optimal  $k$  for the training half was then sought with respect to the validation half. Such a random split followed with a  $k$ -NN learning session was performed  $L$  times to obtain  $L$  values:  $k_1, \dots, k_L$ . The learning routine is also presented in Fig. 3 in a pseudo-Pascal code.

---

```

{ Tr – training set }
for i:=1 to L do { L trials }
begin
  RS(i) := random_select(size(Tr) / 2);
  { RS(i) – random subset of Tr }
  CVS(i) := Tr \ RS(i); { CVS(i) – current validation set }
  k(i):=find_best_k_for_kNN(RS(i), CVS(i));
  { on RS(i), with respect to CVS(i) }
end;
{ output: learnt parameters k(1), ..., k(L) }

```

---

**Figure 3: Finding parameters for voting  $k$ -NN**

The classification of a query sample  $q$  consists in producing  $L$  class labels for  $q$  according to  $k_i$ -NN,  $i=1..L$ , and finally assigning  $q$  to the class most frequently appearing among those  $L$  labels. By analogy we construct voting  $k$ -NCN and voting  $k$ -NSN rules.

#### 4 Computational issues

What is the time complexity of finding  $k$  nearest neighbors? In practice, usually a *naive* implementation is used. During the search a sorted list of  $k$  NN's for a query sample is kept and updated when needed. Although typically (and for small enough values of  $k$ ) the search time is about  $1.1n..1.2n$  (i.e. only some 10-20% longer than 1-NN search time), in worst case it will be  $O(d \cdot n \cdot k)$ ,  $d$  – dimension. To protect from the worst case, another

implementation can be used. First, distances to all  $n$  samples from the reference set are calculated and sorted, and then voting over classes of the first  $k$  neighbors incurs the classification decision. The classification time is therefore  $O(d \cdot n + n \log n + k + c)$ ,  $d$  – dimensionality,  $c$  – number of classes. As  $k \leq n$  and  $c \leq n$  (typically  $k \ll n$  and  $c \ll n$ ), and usually  $\log n$  is on the order of magnitude of  $d$ , the overall cost is close to  $O(d \cdot n)$ .

The classification in voting  $k$ -NN may follow the latter of the described implementations. The difference is in the final stage: instead of scanning  $k$  neighbors, we have to take into account  $k_{\max} := \max_{i=1..L} k_i$  neighbors. The overall cost is  $O(d \cdot n + n \log n + k_{\max} + L \cdot c)$ . For reasonable values of  $L$  (in our tests  $L=10$  seemed fairly good) the cost is practically comparable to the cost of original  $k$ -NN in worst-case protecting version. If the naïve neighbor search is used, then voting  $k$ -NN is slightly slower than plain  $k$ -NN, because  $k_{\max}$  is often about twice greater than the globally selected  $k$ .

In the case of voting  $k$ -NCN, the slow-down related to the conventional rule is equal to the ratio of  $k_{\max}$  and the globally selected  $k$ . The slow-down factor for voting  $k$ -NSN is close to an analogous ratio.

## 5 About the data

Our considerations concern images obtained by two sensors installed on an aircraft: a Daedalus 1268 Airborne Thematic Mapper scanner and a fully polarimetric PLC band NASA/JPL airborne radar system. The geographical location was the Feltwell area. The average registration error was on a pixel level. The ATM images were filtered by a linear smoothing and context-sensitive enhancement filter; then they were segmented by a multiband region-growing technique. Five following regions, i.e. classes, were selected: carrots, potatoes, stubble, sugar beet and wheat. Each pixel was described by 9 features, obtained from optical and radar channels. Below a brief feature description is given: features 1 – 6 are responses of the Daedalus sensor for bands from 2 to 7 respectively, feature 7 is a response for the band C with HH polarization, feature 8 is a response of the radar sensor for L band and HV polarization and the feature 9 is a response of the radar sensor for the band P and VV polarization. More detailed description of the data can be found in [18].

Originally, the authors of [18] considered 15 features. However, our experiments have been constrained to the first 9 features out of 15. Our goal is to present rather the new approaches than search for the features, which would enable the smallest error rate. A data set, we deal with, contains 5 classes, 9 features and 5124 pixels.

## 6 Experimental comparison of the classifiers

We conducted experiments on real data taken from a remote sensing application described in the previous section.

Experiments were performed for training sets with sizes of 500, 750 and 1250 samples, each class represented with the same number of instances; the remaining samples formed respective test sets. For each training set size, ten partitions have been made; the presented results are averages over 10 runs. The city-block metric was used in all tests.

The  $k$ 's for plain  $k$ -NN and  $k$ -NCN were found with leave-one-out cross-validation. As mentioned earlier, the  $k$ -NSN experiments benefited from learning sessions of  $k$ -NCN. The  $k$ -NSN based tests were run with the number of mutations for  $k$ -NSN sessions equal to 500 and 2500. All voting algorithms were tested in the described manner with 10 component classifiers. In all trials, both for plain and voting classifiers, the values of  $k$  were inspected in the interval 1..30. The results, in per cent, are presented in Table 1.

		$k$ -NN	voting $k$ -NN	$k$ -NCN	voting $k$ -NCN	$k$ -NSN, 500 mut.	voting $k$ -NSN, 500 mut.	$k$ -NSN, 2500 mut.	voting $k$ -NSN, 2500 mut.
size 500	error (%)	<b>23.1</b>	<b>23.4</b>	<b>23.5</b>	<b>22.7</b>	<b>22.8</b>	<b>22.0</b>	<b>22.9</b>	<b>22.2</b>
	st. dev. (%)	1.0	0.6	1.1	0.7	0.6	0.6	0.8	0.6
size 750	error (%)	<b>22.6</b>	<b>22.2</b>	<b>21.5</b>	<b>20.9</b>	<b>21.1</b>	<b>20.5</b>	<b>21.2</b>	<b>20.5</b>
	st. dev. (%)	0.5	0.6	0.9	0.6	0.7	0.5	0.6	0.5
size 1250	error (%)	<b>21.4</b>	<b>20.7</b>	<b>20.2</b>	<b>19.7</b>	<b>20.0</b>	<b>19.2</b>	<b>19.9</b>	<b>19.2</b>
	st. dev. (%)	0.8	0.9	0.9	0.5	0.8	0.5	0.6	0.4

**Table 1: Comparison of described algorithms on remote sensing data**

It should be noticed that the described voting technique decreased the test errors of all respective base classifiers in all cases, and also decreased the error variance in most cases. The average errors for the most successful classifier (i.e. voting  $k$ -NSN with 500 mutations) were lower than the errors offered by the plain  $k$ -NN rule by more than 2%.

In Table 2 we compare the numbers of NCN neighbors used for each partition of the datasets and the maximal numbers of NCN neighbors in the voting scheme. For example, the ratio of about 1.6 means exactly that on the given dataset the voting algorithm is on the average slower than plain  $k$ -NCN by the factor of about 1.6.

size	classifier case	partition number										mean	ratio $k_{max} /$ global $k$
		01	02	03	04	05	06	07	08	09	10		
500	$k$ in plain $k$ -NCN	11	10	21	14	7	10	15	21	11	17	13.7	1.6
	$k_{max}$ in voting $k$ -NCN	19	26	24	27	25	13	20	19	20	22	21.5	
750	$k$ in plain $k$ -NCN	14	11	11	14	15	10	21	7	20	10	13.3	1.4
	$k_{max}$ in voting $k$ -NCN	17	28	16	29	12	14	20	13	20	17	18.6	
1250	$k$ in plain $k$ -NCN	15	13	11	14	27	7	11	12	14	13	13.7	1.5
	$k_{max}$ in voting $k$ -NCN	22	15	24	18	16	19	21	29	23	20	20.7	

**Table 2: The values of  $k$  selected for each partition**

## 7 Remarks about the $k$ -NN, $k$ -NCN and $k$ -NSN classifiers

The goal of our work was to create a homogeneous ensemble of  $k$ -NN-like classifiers which would differ only in the number  $k$  of neighbors used for prediction by each



component. The ensemble was expected to be less prone for overfitting the training data and an additional motivation was a modest increase of the classification time.

We obtained quite promising results. There is some resemblance between our method and the old idea of weighted  $k$ -NN [19, 20]. Nearer neighbors generally affect more component decisions than farther ones. This is similar (but not equivalent) to setting weights. One difference to the referenced concept must be however stressed: our “weights” are rank-based, not distance-based. Although the pointed effect may shed some light on why the scheme can work, definitely much more insight is required.

In our previous paper [21] we conjectured the introduced voting idea may be applied also to classifiers more complex than  $k$ -NN. Indeed, the results of voting  $k$ -NCN and  $k$ -NSN presented here confirm attractiveness of the approach. The classification speed decrease, really small in the voting  $k$ -NN case, is however greater now, reaching in our experiments about 50% penalty with  $k$ -NCN. This, however, still contrasts with most ensemble methods, where the slow-down compared to a single component classifier is usually proportional to the number of components.

We have not tried to combine the voting idea with any pairwise scheme for multidecision tasks [16, 22]. This is going to be a subject of our future experiments.

A separate contribution of this paper are further tests of the  $k$ -NCN and  $k$ -NSN rules. The surrounding neighborhood concept and in particular its Nearest Centroid realization, poses several interesting questions:

- Does the centroid criterion reflect the neighborhood homogeneity really well? (Note a set of points (=neighbors) on a line could have a gravity center exactly at the test sample  $q$ , but intuitively we would not call this set as lying symmetrically around  $q$ .)
- Even with the centroid criterion in mind, are there possible other efficient methods of optimizing the neighborhood?
- What about voting over several (different) surrounding neighborhoods?
- For a given set, is it possible to predict if a given method succeeds or fails?

Of course, the last question is of much more general importance.

## 8 Reference set size reduction problem

The classifiers for remote sensing problems are usually constructed with the use of the large data sets. For this reason the classification speed may be not satisfactory. The most promising way to make the classification faster consists in reference set reduction. It is an interesting problem how to reduce the reference set or to replace it by a smaller one without a remarkable decrease of the classification quality. Numerous effective reference set reduction algorithms have been devised only for the 1-NN rule that usually yields worse performance as compared to the standard  $k$ -NN rule. For this reason the classical  $k$ -NN rule may first be approximated by the 1-NN rule. To do this, it is sufficient to reclassify the original reference set, i.e. training set, and then to reduce it and use with the 1-NN rule. However, it is also worth to check how behaves the  $k$ -NN rule operating with the reduced sets obtained for the 1-NN rule.

Three most popular reference set reduction [23, 24, 25] algorithms based on consistency idea will be compared with the approach that consists in reference set partitioning. The consistency means that 1-NN rule operating with the reduced set classifies correctly all points from the original reference set. These algorithms determine the size of the reduced set. The procedure based on reference set partitioning described in

[26] can also define the size of the condensed reference set. The term *condensed* is used to stress that the obtained set is not a subset of the primary reference set. A slight modification of this procedure allows to condense the reference set to the desired size. In the next two sections the detailed description of the analyzed methods of the reference set size reduction will be given.

## 9 Reference set reduction algorithms

**Hart's algorithm.** The first point from the reference set is qualified to the (initially empty) reduced reference set. Next, the remaining points of the primary reference set are classified by the 1-NN rule with the current reduced reference set. Each misclassified point is added to the reduced reference set. Such classification of all points from the primary reference set is repeated as long as  $m$  subsequent classifications do not increase the size of the reduced reference set. The first points selected to the reduced reference set can lie far away from the class boundary. This disadvantage of Hart's algorithm has been removed by Gowda-Krishna modification. The Hart algorithm was originally proposed in the paper [23].

**Gowda-Krishna's algorithm.** A mutual distance measure  $mdm(x)$  is associated with each point  $x$  of the primary reference set. The  $mdm(x)$  is calculated in the following way. For the point  $x$  the nearest point  $y$  from the opposite class is found. A number of points from the same class as  $x$  that lie closer to  $y$  than  $x$  to  $y$  is the value of  $mdm(x)$ . Next, all the points of the primary reference set are arranged according to growing values of  $mdm(x)$ . Finally, the Hart algorithm is applied to the reference set ordered in this way. This modification of Hart's algorithm was proposed in the work [24].

**Tomek's algorithm.** Each point  $x$  for which exists a point  $y$ , from another class than  $x$ , such that the internal part of the ball spanned by the points  $x$  and  $y$  does not contain any points from the reference set, is qualified to the reduced reference set. Originally [25], this algorithm was defined in a different way. Furthermore, the two class problem and the Euclidean distance function was assumed.

The authors of the three above described algorithms tried to construct the so called *consistent reduced reference set*, i.e. the set which, when used as the reference set with the 1-NN rule leads to correct classification of all points from the primary reference set. However, in the case of Tomek's algorithm the consistency is not guaranteed.

Gowda-Krishna algorithm produces the smallest size of the reduced set, the Hart's algorithm is the fastest and the Tomek's procedure generates the separating hypersurfaces close to the ones based on the whole reference set.

## 10 Reference set condensation algorithm

To describe the algorithm it will be convenient to introduce the notion of a *diameter* of the set, which is understood as the Euclidean distance between its two farthest points.

**Condensation by multiple reference set partitioning.** At the start point the condensed set contains only one point  $G(1)$ , equal to the reference set gravity center and labeled as the majority of its points. So, the label of this point corresponds to the class most heavily represented in the training set. Then two farthest points  $P_1$  and  $P_2$  of the reference set are found. The first partition is performed by a hyperplane passing in the middle between  $P_1$  and  $P_2$  and orthogonal to straight line that joins these points. The points, which lie closer to the point  $P_1$  than to the point  $P_2$  or in the same distance form

the set  $C(1)$  and the remaining points create the set  $C(2)$ . The previous point  $G(1)$  is replaced by the gravity center of  $C(1)$ . The point  $G(1)$  and the gravity center  $G(2)$  of  $C(2)$  assume the labels as the majority of points in the sets  $C(1)$  and  $C(2)$  respectively. Now, the condensed set contains two points,  $G(1)$  and  $G(2)$ .

Let us assume that the original reference set has been dropped into  $m-1$  subsets  $C(i)$ . The current condensed set contains then  $m-1$  points  $G(i)$ , gravity centers of  $C(i)$ . From among all  $C(i)$  containing at least two points from different classes the set  $C(j)$  with the largest diameter is selected. This set with the help of its two farthest points  $P_1$  and  $P_2$  is divided into two parts  $D_1$  and  $D_2$ . The old  $C(j)$  is replaced by  $D_1$  and the new set  $C(m)=D_2$  is created. Thus, the number of subsets  $C(i)$  has been increased from  $m-1$  to  $m$ . Taking new  $G(j)$  as the gravity center of  $D_1$  and computing  $G(m)$  as the gravity center  $D_2$ , the size of the current condensed set will be increased to  $m$ . In this way, by virtue of the described recursion scheme, the number of subsets can get larger and larger until all  $C(i)$  with at least two points from different classes will be exhausted. So, the size of the condensed reference set is determined. A more formal description of this algorithm can be found in [26].

## 11 Computational results

Five hundred pixels (one hundred from each class) were randomly selected to be used as the training set, and the remaining 4624 pixels were treated as the testing set. Such an experiment was repeated 10 times. In each experiment all the algorithms presented in the previous two sections were applied.

The optimum value of  $k$  was found first for the whole training set by use of the *leave one out* method. If more than one value of  $k$  offered the smallest error rate then largest one was selected since such a choice promises a lower standard deviation of the misclassification rate. Then this  $k$  was used to reclassify all ten training sets. The raw and the reclassified training sets were separately applied for the reduced and condensed reference set constructions. The values of  $k$  for the reduced (condensed) sets were calculated also by the *leave one out* method only on the basis of these sets. Error rates were calculated by use of corresponding testing sets. Results of the computations were gathered in Table 3. The most interesting results have been marked by the bold font.

One can notice that Tomek's algorithm gives very weak reduction. The error rates are nearly the same as the ones offered by the complete reference set, but the reduction degree is very small. The use of 1-NN rule with the reclassified set instead of  $k$ -NN with raw data is fruitful. The former results in 205 points in the reduced set and error rate equal 26.5%, while the latter produces 294 points and the misclassification rate equal 32.9%. No significant difference is observed between the Hart and the Gowda-Krishna procedures.

Remarkably better results, taking into account the reduction degree as well as the error rate, promises the approach based on reclassified reference set dividing. It offers twice smaller reference set size reduction saving nearly the same performance as Tomek's algorithm. The solution with  $k$ -NN rule,  $k>1$ , costs slightly more time in comparison to the case when  $k=1$ . A suitable modification of the well known Quick Sort algorithm allows very fast search of the nearest neighbors.

The result of Hart's and Gowda-Krishna's approaches are also worth of attention because of strong reduction.

Decision rule →	Reference set							
	Original					Reclassified		
	1-NN		k-NN			1-NN		
Reduction type ↓	$m_{red}$	<i>error rate %</i>	$k$	$m_{red}$	<i>error rate %</i>	$m_{red}$	<i>error rate %</i>	
Column	0	1	2	3	4	5	6	7
<b>No reduction, mean value</b>	500	30.0	<b>9</b>	<b>500</b>	<b>23.1</b>	500	26.1	
<i>Standard deviation</i>	0	1.0	2	0	0.8	0	0.6	
<b>Tomek, mean value</b>	473	29.9	8	473	23.5	419	26.1	
<i>Standard deviation</i>	4.9	0.9	3	4.9	1.0	12	0.6	
<b>Hart, mean value</b>	255	34.7	15	255	27.4	<b>162</b>	<b>28.3</b>	
<i>Standard deviation</i>	10	1.0	4	10	2.4	11	1.3	
<b>Gowda-Krishna, mean value</b>	245	35.0	26	245	29.3	<b>149</b>	<b>28.9</b>	
<i>Standard deviation</i>	12	0.6	12	12	2.9	11	1.3	
<b>Partitioning, mean value</b>	294	32.9	<b>11</b>	<b>294</b>	<b>25.1</b>	<b>205</b>	<b>26.5</b>	
<i>Standard deviation</i>	13	1.0	5	13	1.5	14	1.2	

Table 3: Results of reference set size reduction

## 12 Conclusions

As it was explained in Section 1, the classifiers based on  $k$ -NN rule promise the performance close to the one offered by the Bayes classifier if the training sets are sufficiently large. We have shown that some modifications of the  $k$ -NN rules can outperform the original standard version, it depends on the data we deal with. The larger are the training sets, the smaller difference between the standard and the modified version of the NN type classifier can be expected. We feel that also the performance of the  $k$ -NN rules or 1-NN rule operating with the reduced or the condensed sets converge to the performance of the Bayes classifier. Thus, for the very large data there will be no reason to use the reference set of the original size.

In the case of the rules based on the reduced reference sets we have used the standard  $k$ -NN rule with the raw data or 1-NN rule with the reference sets reclassified by the  $k$ -NN rule. It would be interesting to examine the results, which could be obtained by association the reduced reference sets with the proposed  $k$ -NSN classifier. The classifier for remote sensing problems can be based on several time larger data than the ones used in the present paper. The reference set reduction can not only accelerate the classification phase but also the training stage.

It is worth to notice that reference set condensation algorithm starts with one point in the condensed set and stops when each subset of the original reference set contain points from the one class only. So, there is a possibility to control the classification quality after each sequential increasing of condensed set by one point. In this way we can find the most appropriate compromise between the speed and the quality of the classification phase.

## 13 Acknowledgement

The support of NATO Grant PST.CLG.977258 on this work is gratefully acknowledged.

## References

1. Duda R. H., Hart P. E., *Pattern Classification and Scene Analysis*, John Wiley, New York, 1973.
2. Fix E., Hodges J. L., 1952, "Discriminatory Analysis: Nonparametric Discrimination Small Sample Performance," *project 21-49-004, Report Number 11*, USAF School of Aviation Medicine, Randolph Field, Texas, pp. 280–322.
3. Devijver P. A., Kittler J., *Pattern Recognition: A Statistical Approach*, Prentice Hall, London, 1982.
4. Chaudhuri B. B., 1996, "A new definition of neighbourhood of a point in multi-dimensional space," *Pattern Recognition Letters*, Vol. 17, pp. 11–17.
5. Sánchez J. S., Pla F., Ferri F. J., 1997, "On the use of neighbourhood-based non-parametric classifiers," *Pattern Recognition Letters*, Vol. 18, pp. 1179–1186.
6. Sánchez J. S., Pla F., Ferri F. J., 1998, "Improving the  $k$ -NCN classification rule through heuristic modifications," *Pattern Recognition Letters*, Vol. 19, pp. 1165–1170.
7. Grabowski Sz., 2002, "Towards decision rule based on closer symmetric neighborhood," *Biocybernetics and Biomedical Engineering*, accepted.
8. Skalak D. B., 1994, "Prototype and Feature Selection by Sampling and Random Mutation Hill Climbing Algorithms," *Machine Learning: Proceedings of the Eleventh International Conference*, Morgan Kaufmann.
9. Dietterich T. G., 2000, "Ensemble methods in machine learning," *Multiple Classifier Systems. First International Workshop (MCS2000)*, Cagliari, Italy: Springer-Verlag, pp. 1–15.
10. Bauer E., Kohavi R., 1999, "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants," *Machine Learning*, 36, pp. 105–142.
11. Kuncheva L. I., Whitaker C. J., 2002, "Measures of diversity in classifier ensembles," *Machine Learning*, accepted (available at <http://www.bangor.ac.uk/mas00a/papers/lkml.ps.gz>).
12. Shipp C. A., Kuncheva L. I., 2002, "Relationships between combination methods and measures of diversity in combining classifiers," *Information Fusion*, Vol. 3, pp. 135–148.
13. Skalak D. B., 1996, "Prototype Selection for Composite Nearest Neighbor Classifiers," PhD thesis, Dept. of Computer Science, University of Massachusetts.
14. Alpaydin E., 1997, "Voting over multiple condensed nearest neighbors," *Artificial Intelligence Review*, Vol. 11 (1-5), pp. 115–132.
15. Bay S. D., 1998, "Combining Nearest Neighbor Classifiers Through Multiple Feature Subsets," *Proceedings of the Fifteenth International Conference on Machine Learning*, Madison, WI.
16. Jozwik A., Vernazza G., 1988, "Recognition of leucocytes by a parallel  $k$ -NN classifiers," *Lecture Notes of ICB Seminar*, Warsaw, Poland, pp. 138–153.
17. Ricci F., Aha D. W., 1998, "Error-correcting output codes for local learners," *Proceedings of the Tenth European Conference on Machine Learning*, Chemnitz, Germany.
18. Serpico S. B., Roli F., 1995, "Classification of multisensor remote sensing images by structured neural networks," *IEEE Trans. Geoscience Remote Sensing*, Vol. 33, No. 3, pp. 562–578.
19. Dudani S. A., 1976, "The Distance-Weighted  $k$ -Nearest Neighbor Rule," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 6, No. 4, pp. 325–327.

20. Wilson D. R., Martinez T. R., 2000, "An Integrated Instance-Based Learning Algorithm," *Computational Intelligence*, Vol. 16, No. 4, pp. 1–28.
21. Grabowski Sz., 2002, "Voting over Multiple k-NN Classifiers," *Proceedings of the IEEE Conference TCSET'2002*, Lviv-Slavsk, Ukraine, pp. 223–225.
22. Moreira M. Mayoraz E., 1998, "Improved pairwise coupling classification with correcting classifiers," *Proceedings of the Tenth European Conference on Machine Learning*, Chemnitz, Germany.
23. Hart P.E., 1968, "The condensed nearest neighbor rule," *IEEE Trans. Information Theory*, Vol. 14, No. 3 (Corresp.), pp. 515–516.
24. Gowda K. C., Krishna G., 1979, "The condensed nearest neighbor rule using the concept of mutual nearest neighborhood," *IEEE Trans. Information Theory*, Vol. 25, No. 4, pp. 488–490.
25. Tomek I., 1977, "Two modifications of CNN," *IEEE Trans. Systems, Man, and Cybernetics*, Vol. 7, No. 2, pp. 92–94.
26. Chen C. H., Jóźwik A., 1996, "A sample set condensation algorithm for the class sensitive artificial neural network," Vol. 17, pp. 819–826.

This page is intentionally left blank

## CHAPTER 14

# UNSUPERVISED SEGMENTATION OF HYPERSPECTRAL IMAGES USING GAUSS-MARKOV RANDOM FIELDS AND PCA

C. H. Chen and Hao Chen

*University of Massachusetts Dartmouth, N. Dartmouth, MA 02747-2300 USA*  
*E-mail: [cchen@umassd.edu](mailto:cchen@umassd.edu)*

Processing, analysis and transmission of the remote sensing data require large amounts of computation and storage space. Both the Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are useful to reduce the remote sensing data size. Both are globally optimum according to some criteria. This chapter introduces the Gauss-Markov random field, which is assumed to be the model of observed terrain, and the maximum *a posteriori* (MAP) estimation for remote sensing data compression and unsupervised classification. The PCA with MRF method and clustering algorithms are applied to an AVIRIS data (Hyperspectral imagery data). By comparing with the result of PCA/ICA and k-mean algorithm, remote sensing data compressed with PCA and MRF can be more easily classified by the unsupervised classification algorithm. However the popular FCM (Fuzzy-c-means) algorithm does not perform significantly better in the remote sensing data than the k-mean method though it needs larger amount of computation.

### 1. Introduction

The remote sensing imagery data can provide high quality information about observed ground, which can be utilized for background characterization, object recognition etc. The high dimensionality of remote sensing data, such as Hyperspectral Image Data (HSI) with hundreds of spectral bands, promises to provide better background information and other potential applications such as target detection. However, high dimensional data brings new problem in data storage, transmission and analysis. To obtain the information of background characterization of observed region, the supervised or un-supervised classification can be performed according to the data type and attributes. No matter what classification method is used, the reduction of data dimensionality should be conducted first to allow effective feature extraction for classification because the redundancy of remote sensing data will cause unacceptable storage and computation. Reduced feature dimension will also minimize the problem with the Hughes phenomenon in pattern classification [1]. Principal Component Analysis (PCA) is generally used to de-correlate data and maximize the information content in a reduced number of features. This maximization of information is based on the covariance matrix of different spectral bands. But the principal components contain only the background of observed terrain due to its global optimization. Some small targets or small edges are possibly smoothed out.

Markov random field (MRF) theory provides a basis for modeling contextual constraints in signal/image processing, analysis and interpretation [2]. Several methods are introduced and published about 2-D image restoration or reconstruction with MRF model [3-7]. Also, MRF is used in remote sensing data classification [8-11]. In this chapter, we will use MRF model in the remote sensing data for data compression. With



MRF model, not only the correlation of different bands but also the correlation of neighborhood pixels is exploited to reduce remote sensing data size. The observed ground region is assumed to satisfy the MRF model and the maximum *a posteriori* (MAP) estimation is used to compress and construct the remote sensing data with PCA method [12,13]. The results on one Hyperspectral Image Data (HIS) data set will be presented to illustrate the effectiveness of the proposed approach.

## 2. Markov Random Field And PCA

Various vision/image models using MRF have been introduced to solve the practical problems such as image restoration and reconstruction, edge and region segmentation, texture, stereo and motion, object matching and recognition. The Markov Random Field (MRF) is first used to restore/reconstruct noisy image by Geman and Geman [3]. The simulated annealing algorithm is introduced from the Gibbs distribution, which is equivalent to the MRF, and has a good performance to reconstruct the badly degraded images. Based on this work, the compound Gauss Markov random field (CGMRF) introduced by Jeng and Woods[4] provides a faster convergence algorithm compared to the previous ones that have good performance in practical image restoration.

### 2.1 Neighborhood System

Let  $\mathbf{x} = \{ x_{i,j} \in \mathfrak{R} : i=1, 2, \dots, M; j 1, 2, \dots, N \}$ , containing  $M \times N$  real valued pixels, be a sample of the image. Let  $L$  denote the set of the pixels in the image. Each pixel  $x_{i,j}$  in the image are related to others via a neighborhood system. The neighborhood system of  $L$ , denoted as  $\mu = \{ \mu_{i,j} : (i, j) \in L, \mu_{i,j} \subseteq L \}$ , can be defined as any subsets of  $L$ , which satisfy [15],

1. The pixel is not neighboring to itself:  $(i, j) \notin \mu$ ;
2. The neighboring relationship is mutual: if  $(k, l) \in \mu$  and  $\eta$  is the neighborhood system of  $(k, l)$ , then  $(i, j) \in \eta$ .

The neighborhood system with a definitive order is shown in Figure 1 (a), where 1, 2, ..... denote the order of neighborhood system. The second order neighborhood system has eight neighboring pixels as Figure 1 (b). Therefore, it is also called the 8-neighborhood system.

### 2.2 Markov Random Fields

The image  $\mathbf{x}$  is a Markov random field on  $L$  with respect to the neighborhood system  $\eta$  if and only if the following two conditions are satisfied:

1.  $P(\mathbf{x}) > 0, \forall \mathbf{x} \in X$
2.  $P(x_{i,j} | x_{k,l} : (k, l) \in L; (k, l) \neq (i, j)) = P(x_{i,j} | x_{k,l} : (k, l) \in \eta_{i,j})$

Those conditions are introduced for some applicable reasons and can be satisfied in practice. When the second condition is satisfied, the joint probability  $P(\mathbf{x})$  will be uniquely determined by its local conditional probabilities. So the MRF represents the local characteristics of  $\mathbf{x}$ . In other words, only neighborhood pixels have direct interactions with each other. A MRF can have other properties such as homogeneity and isotropy. It is said to be homogeneous if  $P(x_{i,j} | x_{k,l} : (k, l) \in \eta_{i,j}^r)$  is independent of the position of current pixel  $(i, j)$  in  $L$ . The random field satisfying the above condition is called Markov random field with the neighborhood system  $\eta$ .

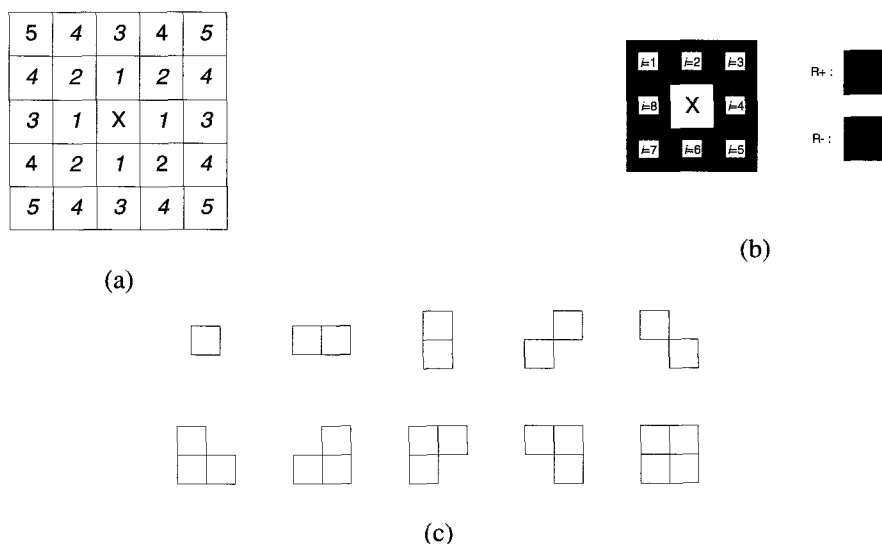


Figure 1 (a) ordered neighborhood system for MRF.

(b) The second order neighborhood (c) The clique types of the second order neighborhood system

### 2.3 Gibbs Random Fields

A set of random variables  $\mathbf{x}$  is said to be a Gibbs random field (GRF) on  $L$  with respect to  $\eta$  if and only if its configurations obey a Gibbs distribution. A Gibbs distribution takes the following form:

$$P(\mathbf{x}) = \frac{1}{Z} \exp\left(-\frac{U(\mathbf{x})}{T}\right) \quad (1)$$

where  $Z$  is a normalizing constant called the partition function,  $T$  is a constant parameter called the temperature which is usually assumed to be 1 unless otherwise stated, and  $U(\mathbf{x})$  is the energy function [15][16], which is the sum of clique potentials. The cliques of the image are defined to be the subset of the  $L$ . The clique types of the second-order neighborhood system are shown in Figure 1 (c).

It is said to be isotropic if the clique potentials are independent of the orientation of clique. The isotropy is a property of direction independent regions. To assume the data homogeneous and isotropic will considerably simplify the model description and reduce the complexity of algorithm. The MRF is characterized by its local property (the Markov chain) whereas a GRF is characterized by its global property (the Gibbs distribution). By the Hammersley-Clifford theorem, these two types of random fields are equivalent. In this chapter, the remote sensing data is assumed homogeneous and isotropic.

### 2.4 Principal Component Analysis (PCA)

The PCA (*Karhunen-Loeve transform*) is a linear transformation for a sample in  $n$ -dimensional space which makes the new coordinate axes uncorrelated. So PCA method has in practice been used to reduce the dimensionality of problems, and to transform interdependent coordinates into significant and uncorrelated ones. Let  $\mathbf{x}$  denote an  $n$ -dimensional random vector and assume it has zero mean,  $E(\mathbf{x}) = 0$  where  $E(\cdot)$  is statistical expectation. The covariance matrix  $C_{xx}$  of the sample  $\mathbf{x}$  is defined as:

$$C_{xx} = E(\mathbf{x} \cdot \mathbf{x}^T) \quad (2)$$

Let  $\lambda_1, \lambda_2, \dots, \lambda_n$  be the eigenvalues of the matrix  $C_{xx}$  and  $\lambda_1 > \lambda_2 > \dots > \lambda_n$ . Also the matrix  $V = [v_1, v_2, \dots, v_m]$  ( $m \leq n$ ) is constructed by the eigenvectors  $v_1, v_2, \dots, v_m$  of  $C_{xx}$ , where  $v_1, v_2, \dots, v_m$  are associated with  $\lambda_1, \lambda_2, \dots, \lambda_m$ , respectively. The matrix  $V$  is orthogonal and its column vectors satisfy the orthonormal condition. We project  $x$  onto the principal directional axes by matrix  $V$ :

$$z = V^T x; z_j = v_j x \quad j=1, 2, \dots, m; m \leq n \tag{3}$$

Here  $z_j$  is the projections of  $x$  onto the  $j$ -th principal directional axis and called the  $j$ -th principal component. When  $m \leq n$ , data dimension is reduced in the minimal mean square error sense, the error being equal to the sum of the  $\lambda_{m+1}, \lambda_{m+2}, \dots, \lambda_n$ . So it can be said that the PCA method uses second order statistics information to reduce the data dimension in the minimal mean square error sense.

### 3. Gaussian MRF for Remote Sensing Data Compression

With good quality remote sensing data such as SAR data, multi-spectral imagery data etc, the detail of a ground region can be obtained. In general, data compression is necessary to extract desired information from remote sensing data especially for HSI data, which consisted of hundreds of spectral bands. In this section, the compound Gaussian Markov field will be used to compress the remote sensing data with PCA method.

#### 3.1 Compound Gaussian MRF Data Model [4]

Under the homogeneous MRF assumption of image model with respect to a certain neighborhood system  $\eta$ , in our experiments the second-order neighborhood system is chosen, any pixel  $x_{i,j}$  of image can be predicted by the linear combination of the its neighborhood system:

$$x_{i,j} = \sum_{k,l \in \eta} c_{i-k,j-l} x_{k,l} + n_{i,j} \tag{4}$$

where the parameter  $c_{i-k,j-l}$  is independent of the position of pixel  $(i,j)$  if the model is homogeneous and  $n_{i,j}$  is a Gaussian random field. The conditional probability function of pixel  $x_{i,j}$  given its neighboring system is a Gaussian distribution:

$$p(x_{i,j} | \eta_{i,j}) = \frac{1}{\sqrt{2\pi\sigma_n}} \exp \left[ -\frac{\left( x_{i,j} - \sum_{k,l \in \eta} c_{i-k,j-l} x_{k,l} \right)^2}{2\sigma_n^2} \right] \tag{5}$$

The image sample is arranged initially into a column vector  $x = [x_1, x_2, \dots, x_{M \times N}]^T$  in the lexicographic order.  $M$  and  $N$  are row size and column size of image data respectively. This column vector will then be a set of jointly Gaussian random variables that also possess the Markov property. Thus the joint probability density function of the random variables constituted by the pixel values in  $x$  has the form:

$$p(x) = \frac{|A|^{\frac{1}{2}}}{(2\pi)^{MN/2}} \exp \left\{ -\frac{1}{2} x^T A x \right\} \tag{6}$$

where  $A$  is the inverse of the covariance matrix of the random variable  $x$ .  $|A|$  is its determinant. Since  $x$  can be represented by a GMRF model, the conditional probability density function will be only dependent on its neighbor  $\eta$ .

As proposed by Jeng and Woods [4], the relationship of a pixel in compound GMRF and its neighborhood system is presented by the collection of binary variables  $C = \{c_{m,n}^k : k = 1, 2, \dots, 8; m = 1, 2, \dots, M; n = 1, 2, \dots, N\}$ , where  $k$  is the same as the Figure 1 (b). If the neighbor pixels are independent, the associated  $c_{m,n}^k$  is set to be one and otherwise  $c_{m,n}^k$  is equal to zero. The conditional probability density function of  $x$  given the edge configuration  $C$  is given by:

$$p(x|C) = \frac{1}{Z_1(C)} \exp \left\{ -\frac{\mu}{2} \sum_{n=1}^N \sum_{m=1}^M \left[ \sum_{k=1}^8 c_{n,m}^k w_k (x_{n,m} - x_{n-k,m-k})^2 + \left( 1 - 2 \sum_{k=1}^8 w_k \right) x_{n,m}^2 \right] \right\} \quad (7)$$

where  $\mu$  is the “global smoothness” of the compound GMRF, the inverse of the “global variance”. The  $w_1, w_2, \dots, w_8$  control its relative “smoothness” of eight directions which correspond to eight neighborhood pixels in the second order neighbor system. In general, the MRF model for image data can be assumed as homogeneous. So the  $w_i$  is independent of the position of the pixel in the image.  $Z_1(C)$  is the normalizing constant called partition function. We can rewrite equation (7) in vector form as:

$$p(x|C) = \frac{[\mu A(C)]^{\frac{1}{2}}}{(2\pi)^{MN/2}} \exp \left\{ -\frac{\mu}{2} x^T A(C) x \right\} \quad (8)$$

### 3.2 Observed Data and Maximum a Posteriori (MAP) Estimation

Each observed data  $y$  can be regarded as the noisy version of original data  $x$ . we can write  $y = x + n$ , where  $n$  is a sample of a white Gaussian noise field with variance  $\sigma_n^2$ . The conditional probability density function of  $y$  given  $x$  can be easily given by:

$$p(y|x) = \frac{1}{(2\pi\sigma_n^2)^{-MN/2}} \exp \left\{ -\frac{1}{2\sigma_n^2} (y-x)^T (y-x) \right\} \quad (9)$$

The estimation problem is to obtain the original data  $x$  from the observed data. With the compound MRF image model of  $x$  in the previous section, the estimation of original image  $x$  would be acquired with observed  $y$  and the matrix  $A(C)$ , which should be estimated/obtained from the observed  $y$ . In general, we have  $p(x|C,y) \propto p(x|C)p(y|x)$ :

$$p(x|C,y) \propto \frac{1}{(2\pi\sigma_n^2)^{-MN/2} * Z_2(C)} \exp \left\{ -\frac{1}{2\sigma_x^2} x^T A(C) x - \frac{1}{2\sigma_n^2} (y-x)^T (y-x) \right\} \quad (10)$$

where  $\sigma_x^2$  represents a “global variance” of the image.

Maximizing the equation should satisfy:

$$\hat{x}_{MAP} = \left( I + \frac{\sigma_n^2}{\sigma_x^2} A(C) \right)^{-1} y = \left( I + kA(C) \right)^{-1} y \quad (11)$$

where  $I$  is the identity matrix and the variance ratio  $k$  has to be estimated.

### 3.3 Compound MRF and PCA for Remote Sensing Data

Let  $Y$  denote the remote sensing data set containing  $L$  dimensionality which can be composed of different spectral bands or polarized directions. The size of observed region is  $M \times N$  pixels which are real valued pixels. With the PCA method, we can reduce the data size with equation (3).

With PCA method, the remote sensing data is projected onto the  $m$  principal directions and the data dimensionality is reduced. Because maximizing information is based on the covariance matrix of different spectral bands, this method is just the global optimization and does not keep the small target or small edge information well, which is mainly the local information.

The introduction of the compound MRF model to compress the remote sensing data will keep the small features of a terrain. Each image of the remote sensing data can be considered as one noisy version of observed terrain. Let  $Z$  be the  $L$  independent information sources of terrain and  $N$  the noise vector introduced from the total sensor system. The remote sensing data can be presented as:

$$Y = U*Z + N \quad (12)$$

where each image data is arranged initially into a row vector of  $Y$  in the lexicographic order. The data  $Z$  is what we wanted. With the description in previous section, the final estimation with reduced data dimension can be written as:

$$\hat{Z} = V\hat{X} = VY \left( (I + kA(C))^{-1} \right)^T \quad (13)$$

where  $V$  is defined in equation (3). Applying equation (13) to reducing dimensionality of the estimating data  $\hat{Z}$ , the parameter  $k$  and matrix  $A(C)$  play a fundamental role in the problem at hand.

### 3.4 Parameter Estimation and Selection

The parameter estimation and selection are important for the MRF data model [17]. In general, the Maximum likelihood estimator (MLE) [5], Least Squares estimator and Minimum Description Length Principle [6] can be applied in the parameter estimation in a single image. Because of its high dimensionality of the remote sensing data, a distinct method should be used to estimate the parameter.

#### 3.4.1 The matrix $A(C)$

The  $A(C)$  is dependent on the parameter  $w_i$  and  $c_{n,m}^i$  ( $n=1, 2, \dots, N$ ;  $m=1, 2, \dots, M$  and  $i=1, 2, \dots, 8$ ). The parameter  $c_{n,m}^i$  is the binary value with edge information. Each of those, when set to zero, breaks the direct bond between pixels in the clique. The  $C$  should be obtained through parameter estimation. With a large data dimensionality of remote sensing data,  $C$  can be obtained from the correlation of pixel pairs within the neighborhood system.

Let  $y$  denote a  $K$  dimensional random vector, which can be a column vector of remote sensing data  $y$  and assume it has zero mean. The cosine of the angle between two vectors is defined by:

$$\cos(\theta(y_1, y_2)) = \frac{\langle y_1, y_2 \rangle}{\|y_1\|_2 \|y_2\|_2} \quad (14)$$

where  $\langle x, y \rangle$  is the Euclidean inner product ( $\langle x, y \rangle = x^T y$ ); and  $\|x\|_2$  is the Euclidean norm of vector  $x$  ( $\|x\|_2 = \langle x, x \rangle^{1/2}$ ). Using the Cauchy-Schwarz inequality, the cosine of any two vectors will be over  $[-1, 1]$  interval. So we can set the parameter  $c_{n,m}^i$  can be set as:

$$c_{n,m}^i = \begin{cases} 1 & \cos(\theta(y_{n,m}, y_{n,m}^i)) \geq \alpha \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

where  $y_{n,m}^i$  is the  $i$ -th neighbor of the pixel  $y_{n,m}$  (see Figure 1),  $\alpha$  is the threshold and can set between  $[-1, 1]$ . The  $\alpha$  is decided according to the smoothness of remote sensing data. The smoother the data the larger the  $\alpha$  is.

The condition  $1 - 2 \sum_{k=1}^8 w_k > 0$  is sufficient for  $A(C)$  to be nonsingular for any  $c_{n,m}^i$

[4]. The  $w_i$ , ( $i = 1, 2, \dots, 8$ ) denotes the relative “smoothness” of its direction. In usual and practical situations, the MRF model is isotropic for image data. So all  $w_i$  will be equal. This indicates the ‘smoothness’ is independent of the directions. Thus the condition can be simplified as  $w < 1/16$ . We set the  $w$  equal to 0.062 for the all experiments in this paper.

### 3.4.2 Coefficient $k$

With the equation (11), we can see that the coefficient  $k$  is the ratio of noise variance and “global variance” of image data and should be estimated with the original data. The noise variance can be obtained from the estimated value and observed data by:

$$\sigma_n^2 = \frac{1}{MN} \sum_{n=1}^N \sum_{m=1}^M (\hat{x}_{n,m} - y_{n,m})^2 \quad (16)$$

Because each remote sensing data set is only a sample of observed region, the exact “global variance” cannot be obtained with just one data set. Figueiredo uses the minimum description length to estimate the global variance of 2D image [6]. The remote sensing data can be regarded as data set of several 2D image data. For hyperspectral image data, there are hundreds spectral bands. Estimating the global variance with MDL will mean huge amount of computation. At the same time, the remote sensing data have abundant information. So we can consider using other method to estimate this parameter. Although the exact equation for global variance couldn’t be obtained, we can say that the global variance of remote sensing data with Gaussian MRF model can satisfy following relation:

$$\sigma_x^2 \propto \frac{1}{NM} \sum \sum \sum c_{n,m}^i (\hat{x}_{n,m} - \tilde{x}_{n,m}^i)^2 \quad (17)$$

where  $c_{n,m}^i$  is the same in the equation (15),  $\hat{x}_{n,m}$  is an estimate of current pixel. With this relation, the coefficient  $k$  will satisfy the relation:

$$k \propto \frac{\sum_{n=1}^N \sum_{m=1}^M (\hat{x}_{n,m} - y_{n,m})^2}{\frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M \sum_{i=1}^8 c_{n,m}^i (\hat{x}_{n,m} - \tilde{x}_{n,m}^i)^2} \quad (18)$$

The coefficient  $k$  is belonging to the region  $[0, 1]$ . This can be satisfied in most situations because the noise power is less than the image data power. So the determinant of  $kA(C)$  will be much smaller than one especially with the large size image data because only the position near the diagonal of matrix is non-zero and the each element of the matrix is not greater than one.

In practice, when  $k$  is equal to the real ratio of  $\sigma_n^2$  and  $\sigma_x^2$ ,  $k$  will be stable in its neighbor region. It’s said that the right side of (18) should be local maximum or minimum. In the previous description,  $k$  is selected in the region  $[0, 1]$ . So, we can determine the coefficient  $k$  with the iteration: initializing the  $k$  and increasing  $k$  with a variable step, which will decrease slowly according to the estimation error, until the left side of (19) is stable. Figure 2 (left) shows the  $k$  taking a maximum value over  $[0, 1]$ .

The simulation result is presented as follows. The parameter  $\alpha$ , which is located in equation (15), is equal to 0.4. The white Gaussian noise is added to the remote sensing data and estimated by previous method. The standard deviation of noise is 5, 6, 7 and 8 respectively. The estimates of  $\sigma_n$  approach the real values when the iteration increases. Table 1 shows the final estimates of  $\sigma_n$ . The error of estimates is below 3%, which can be considered as reasonable.

$\sigma_n$	5.00	6.00	7.00	8.00
Estimate	5.09	6.00	6.90	7.82
Estimation error (%)	1.8	0	1.4	2.2

Table 1 The Std of noise and its estimation

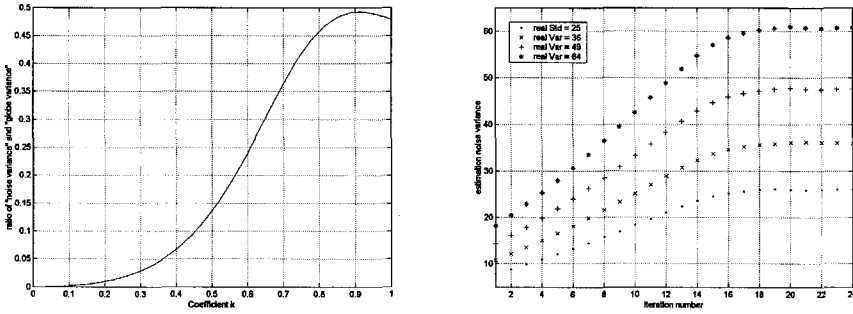


Figure 2 (left) the estimated ratio of “noise variance” and “global variance” relative to the coefficient  $k$ ; (right) The estimation of  $\sigma_n$  and iteration number

### 3.5 Simplifying the Algorithm

There is a matrix inverse operation in equation (13). When the size of observed region increases with  $n$ , the size of matrix  $A$  will increase with  $n^2$ . For example, the matrix  $A$  will be  $10,000 \times 10,000$  when the data size is just  $100 \times 100$ . So the inverse operation of such matrix will need huge computation unless most elements in the matrix  $A$  are zeros. Because only the elements near the diagonal of  $A$  are nonzero and their values are much smaller than 1 ( $w_i$  is equal to 0.062) and coefficient  $k$  is less than one, the determinant of  $kA$  is greater less than one especially when the data size is great. So we can simplify Equation (11)

$$\left( (kA(C) + I)^{-1} \right)^T = \left( I + \sum_{i=1}^{\infty} (-kA(C))^i \right)^T \approx I - kA(C) \tag{19}$$

Note also that the matrix  $A(C)$  is symmetrical because neighborhood system is mutual:

$$A(C)^T = A(C) \tag{20}$$

With this simplified equation (21), we always can obtain the estimate  $\hat{X}$  from the HSI data  $Y$  whether the  $(I + kA(C))$  is singular or nonsingular.

### 3.6 Algorithm Structure

Equation (13) presents the basic idea on how to compress the remote sensing data with compound MRF and PCA method. Several parameters should be estimated before applying this equation. The total algorithm is carried out in the following steps:

1. Estimate the potential matrix  $A(C)$  from remote sensing data  $Y$  with equation (15);
2. Estimate the coefficient  $k$  given  $A(C)$  and remote sensing data  $Y$ ;
3. Get the transform matrix  $V$  with PCA method;
4. Obtain  $\hat{Z}$  with the reduced data dimensionality.

This algorithm includes one loop which is basically an iterative parameter estimation scheme. The complete structure of the algorithm is depicted in Figure 3.

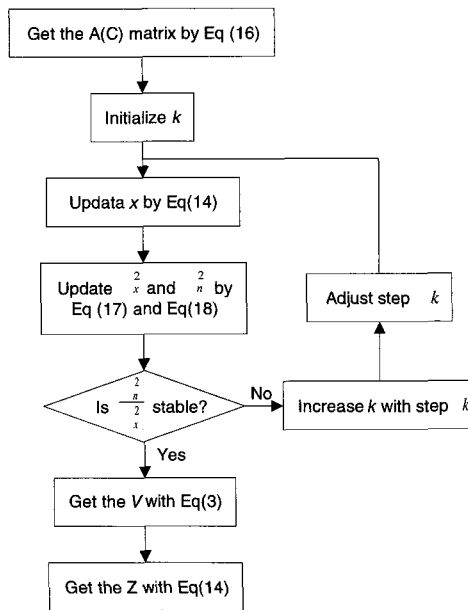


Figure 3 The structure of the proposed algorithm

## 4. FCM and its Extension

### 4.1 Fuzzy-C-Mean Algorithm (FCM)

Clustering is the process to group a sample set of feature vectors into  $K$  cluster sets via an appropriate similarity criterion in the self-organizing mode. The  $k$ -mean algorithm and the ISODATA algorithm are two common clustering algorithms which use classical (crisp) logic. An element either belongs or does not belong to a set. In a fuzzy set, however, the elements can have the partial membership in multiple classes, which has a value anywhere between 0 and 1. The membership grade is associated with each element to indicate the degree to which the element belongs to the set.

The fuzzy  $c$ -means (FCM) [18] algorithm also is an iterative clustering method for unsupervised classification of the cluster data set. Given a data set of samples  $Y = \{y_1, y_2, \dots, y_N\}$  and each element  $y_i$  is a cluster with  $m$  real values. This data set should be classified into  $c$  class. So the objective of FCM segmentation is to generate the cluster centers  $V = \{v_1, v_2, \dots, v_c\}$  and the class membership matrix. The class membership matrix, also called fuzzy partition, has the form  $U = \{u_1, u_2, \dots, u_N\}$ . Each element  $u_i = [u_{i,1}, u_{i,2}, \dots, u_{i,c}]^T$  is a  $c$ -dimensional vector with real value, in which  $u_{i,j}$  represents that  $y_i$  belonging to the class  $j$ , whose center cluster is  $v_j$ , with the membership grade  $u_{i,j}$ . Also the fuzzy partition of  $Y$  is defined as:

$$\{ U \in \mathfrak{R}^{c,N} : u_{i,j} \in [0, 1] \}; \sum_{j=1}^c u_{i,j} = 1 \text{ for any } i; \quad (21)$$

The condition  $\sum_{j=1}^c u_{i,j} = 1$  means that the sum of fuzzy partition should be equal to 1

for each sample cluster. The cost function  $J_m$  of FCM algorithm is defined as a function which projects the fuzzy partition matrix into the real interval  $[0, \infty]$ . That is  $J_m$  is defined as:



$$Jm(U, v) = \sum_{k=1}^N \sum_{i=1}^c (u_{k,i})^m \|y_k - v_i\|^2 \quad (22)$$

where  $\|y_k - v_i\|^2$  are square of the Euclidean distance between  $y_k$  and  $v_i$ . Therefore, the task of algorithm becomes minimizing of  $Jm$  with weighting exponent  $m$ . This leads to the well known FCM clustering algorithm introduced by Bezdek in 1984. The weighting exponent  $m$  has the effect of reducing the squared distance error by an amount that depends on the membership grade in the cluster. As  $m \rightarrow 1$ , the partitions that minimize  $Jm$  become increasingly hard. Conversely, higher values of  $m$  tend to soften a samples cluster membership, and the partition becomes increasingly blurred. Generally  $m$  must be selected by experimental means. The  $m$  is equal to 3 in all experiments of this paper.

The number of clusters  $c$  is an important parameter for FCM algorithm, which should be determined beforehand. The appropriate number of clusters depends on modeling goals, required accuracy and the distribution of data. For example, the number of clusters should be two when the FCM algorithm is used to make binary decision.

## 4.2 Extended FCM with Cluster Merging and Splitting

In applying FCM algorithm, the cluster number  $k$  and initial cluster center play important roles. Their initial values should be determined beforehand and will affect the performance of the algorithm. Cluster merging and/or cluster splitting are often needed to determine the final set of clusters.

### 4.2.1 Cluster Merging

The cluster merging is based on the similarity of two clusters. The cluster similarity can be defined through several methods [19]. In this section, one new definition will be given out. If the two cluster centers are very close, the fuzzy partition for those cluster centers will be very similar. So the cluster similarity coefficient between cluster center  $v_i$  and cluster center  $v_j$  can be defined by:

$$I_{i,j} = \frac{1}{N} \sum_{k=1}^N |u_{k,i} - u_{k,j}| \quad (23)$$

The similarity between clusters is evaluated pair wise and the most similar pair is merged according to the previous cluster similarity coefficient and a threshold  $\lambda$ . Two clusters  $v_i$  and cluster  $v_j$  are merged simply by adding up the row  $i$  and  $j$  of the fuzzy partition  $U$  and remove the two original rows.

$$u_{k,l} = u_{k,i} + u_{k,j} \quad k = 1, 2, \dots, N \quad (24)$$

where  $l$  is the fuzzy partition of new cluster.

### 4.2.2 Cluster Splitting

When the number of clusters is less than the low limit, which is set before application, cluster splitting is performed to guaranteed number of clusters. Set each cluster with  $m$  real values. The cluster  $l$  with largest the class radius will be split into two clusters by following step:

1. Get the  $m-1$  nearest clusters from the cluster  $l$  and the difference vector  $v_l - v_i$  ( $i = 1, 2, \dots, m-1$ ) between the cluster  $l$  and those clusters;
2. Find the unit vector  $w$  which is vertical to all the difference vector  $v_l - v_i$  ( $i = 1, 2, \dots, m-1$ );
3. Two new cluster centers are equal to  $(v_l + r \cdot w/2)$  and  $(v_l - r \cdot w/2)$ , where  $r$  is class radius obtained by the following equation:

$$r = \frac{\sum_{k=1}^N u_{k,l} \cdot \|v_l - y_k\|}{\sum_{k=1}^N u_{k,l}} \quad (25)$$

where  $\|\cdot\|$  the Euclidean norm of vector.

### 4.2.3 Extended FCM Algorithm

The FCM algorithm with the cluster merging is carried out in the following steps:

1. Set the upper limit to the number of clusters; initialize the cluster centers  $V$  according to this limit and set  $t = 0$ ;
2. If  $t \neq 0$ , calculate the  $c$  cluster centers  $V^{(t)}$  with fuzzy partition matrix  $U^{(t)}$  :

$$v_i^{(t)} = \frac{\sum_{k=1}^N (u_{k,i}^{(t)})^m y_k}{\sum_{k=1}^N (u_{k,i}^{(t)})^m} \quad i = 1, 2, \dots, c \quad (26)$$

3. Update the fuzzy partition matrix  $U(t)$ :

If  $y_k \neq v_i$  for any  $i$  ( $i = 1, 2, \dots, c$ ), then:

$$u_{k,i}^{(t+1)} = \left( (d_{k,i})^{\frac{2}{m-1}} \sum_{j=1}^c \left( \frac{1}{d_{k,j}} \right)^{\frac{2}{m-1}} \right)^{-1} \quad i = 1, 2, \dots, c \quad (27)$$

Else, set:

$$u_{k,j} = \begin{cases} 1 & y_k = v_i \\ 0 & y_k \neq v_i \end{cases} \quad (28)$$

4. Compare  $U^{(t+1)}$  and  $U^{(t)}$

$$\Delta U = \sum_{i=1}^N \sum_{j=1}^c |u_{i,j}^{(t+1)} - u_{i,j}^{(t)}| \quad (29)$$

If  $\Delta U < \varepsilon$ , continue;

Otherwise, set  $t = t+1$  and go to step 2

5. Compare the similarity coefficient  $k$  and the similarity threshold  $\lambda$

If  $k < \lambda$ , merge the similar clusters and go to step 2;

Otherwise, continue

6. If the class number is less than the lower limit, split the largest cluster;

Otherwise, stop iteration and output the fuzzy partition matrix.

The FCM algorithm with cluster merging and splitting can be considered as the extended FCM. So the total algorithm is carried out according to Figure 4.

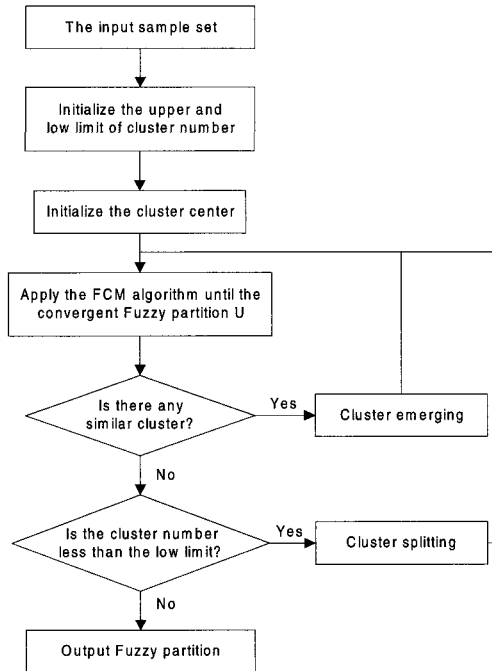


Figure 4 Structure of the extended FCM algorithm

## 5. Application to Hyperspectral Images

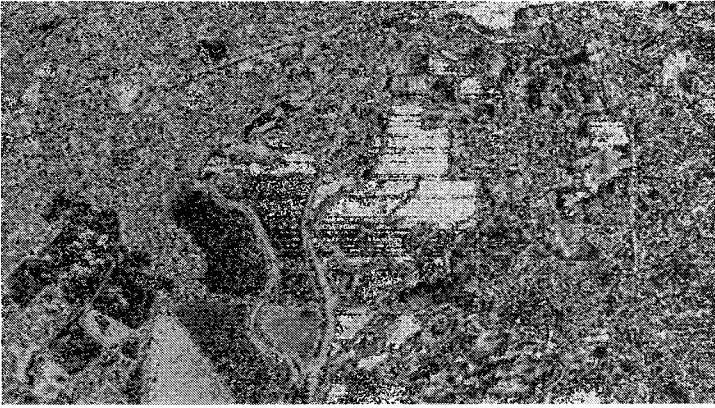
### 5.1 HIS Imagery Data

Hyperspectral Image Data (HSI), with hundreds of high resolution spectral bands, is usually utilized for background characterization. No matter what classification method is used, the reduction of HSI data dimensionality is first conducted to allow effective

AVIRIS stands for the Airborne Visible Infrared Imaging Spectrometer. That is an optical sensor to deliver calibrated images of the upwelling spectral radiance in 224 contiguous spectral channels (bands). The AVIRIS instruments contain 224 different detectors, each with a wavelength sensitive range of approximately 10 nanometers (nm). AVIRIS data used in this experiment is AVIRIS flight number f970620t01p02\_r03, over Moffett Field, California. Image size for each channel is 614\*1024 pixels. We only select one region with size 360\*560 for experiments in this paper. Three images from three spectral bands of the data base are shown in Fig. 5. The truth information about ground cover is unavailable. So we cannot use the quantitative measurements to evaluate the result.

### 5.2 HSI Data Compression with PCA+MRF

Here, we use PCA+MRF to reduce the HSI data size. Figure 6 is first component with the PCA+MRF method. The first component gives the global information about ground cover. The ground cover can be main divided into two types: architecture (dark) region and bare land /plant (bright) region. The following components (Figure 7 and Figure 8 ) give more information about small feature, such as road, building etc. The gray contrast of the fourth component becomes less compared with previous three components. So we just select three components for k-mean algorithm to obtain the ground cover information.



Spectral Band 4



Spectral Band 143



Spectral Band 207

Figure 5



Figure 6 The first component of PCA+MRF method



Figure 7 The second component of PCA+MRF method



Figure 8 The third component of PCA+MRF method

### 5.3 Unsupervised Classification with K-mean Algorithm

From the experimental results on the hyperspectral image data as well as SAR image data, we have determined that the FCM algorithm may not be better than k-mean algorithm in remote sensing data classification. So we use the k-mean algorithm to do unsupervised classification for HSI data. The first three components are used for classification and the class number  $k$  is 8. The result is shown in Figure 10. The bare land/plant region and architecture region in the HSI data can be easily distinguished in Figure 10. However, some small features, such as building blocks, roads, are blurred and hard to identify. Just several wide roads can be seen in this figure. To get more detail about ground cover, the entire method should be improved.

Because PCA is the algorithm to achieve a global minimal mean square error, the first component will contain the most global information. At the same time, the first component has predominated signal power compared with the second and third components. The power of the first components is 31 times of that of second one and 27 times of that of third one for this HSI data set. When we use the first three components to classify the ground cover, the first component will be predominant. So the result will be mostly about global feature of observed region. To avoid this problem, we should equalize the power of those components. In this chapter, pre-classification is introduced before data unsupervised classification. The structure is shown in Figure 11. Each component goes through a k-mean classification and is segmented into 8 classes. With those processing, the power of the three components is comparable. The data is then classified by the k-mean algorithm. The final result is shown in Figure 12. The ground cover contains mainly two classes: architecture region (dark) and bare land/plant region (bright). The road and building block in the architecture region is very distinguished and can be easily identified.

It can be said that the PCA+MRF method with k-mean algorithm can provide good ground cover information with the remote sensing data.

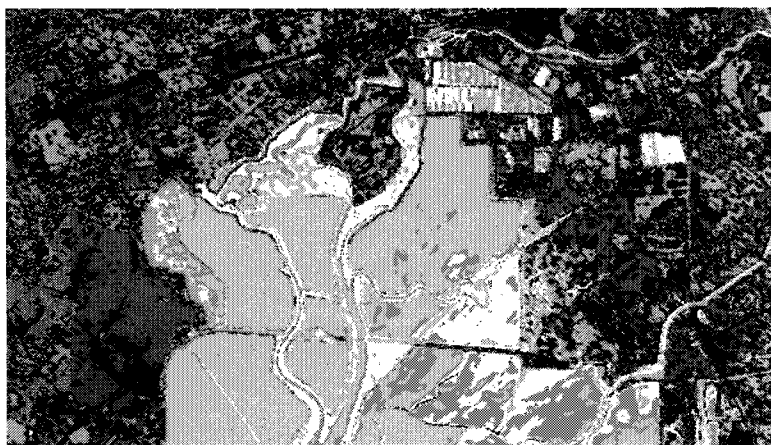


Figure 9 The 8 classes from three components with k-mean algorithm

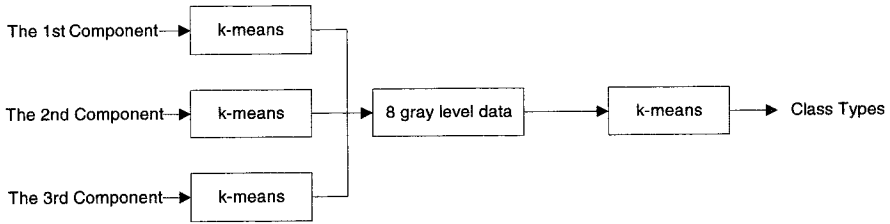


Figure 10 The structure for the improved classification with k-mean algorithm`



Figure 11 The 8 classes from three components with improved algorithm

## 6. Conclusion

In remote sensing data processing and analysis, the interpretation of pixels' relation is complicated. In this chapter, we have developed a new approach using MRF model in remote sensing image data compression. The new method presented is to compress remote sensing data according to the MRF model and PCA method. The method has been able to provide large between-class-variation and thus a good unsupervised segmentation performance. Also, the chapter reviews the FCM algorithm and presents its extension with cluster merging and splitting. However, for the remote sensing data considered the popular FCM algorithm does not provide better performance than the simpler k-mean algorithm. With a large amount of computation required, the FCM algorithm thus is not recommended for remote sensing data processing and analysis.

## References

1. Landgrebe, D. "Information Extraction Principles and Methods for Multispectral and Hyperspectral Image Data", in Chapter 1 of "Information Processing for Remote Sensing", edited by C.H. Chen, World Scientific Publishing 1999.
2. Jiang, M "Mathematical Models in Computer Vision and Image Processing", <http://ct.radiology.uiowa.edu/~jiangm/courses/mm-cv-ip/vision.html>

3. Geman, S. and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721-741, 1984
4. Jeng, F.C. and J.W. Woods, "Compound Gauss-Markov Random Fields for Image Estimation" *IEEE Trans. Signal Processing*, vol. 39, No.3, pp.683-697, 1991
5. Lee, G. G. "On Multiresolution Wavelet Analysis Using Gaussian Markov Random Field", Ph.D dissertation, June 1997, University of Massachusetts-Dartmouth
6. Figueiredo, M.A.T. "Unsupervised Image Restoration and Edge Location Using Compound Gauss-Markov Random Fields and the MDL Principle", *IEEE Trans. Image Processing*, vol. 6, No.8, pp.1089-1101, 1997
7. Bouman, C.A. and M. Shapiro, "A Multiscale Random Field Model for Bayesian Image Segmentation", <http://dynamo.ecn.purdue.edu/~bouman/publications/pdf/>
8. Crawford, M.M and Ricard, M. "Hierarchical Classification of SAR Data Using a Markov Random Field Model", *Southwest Symposium on Image Processing*, Tucson, AZ, pp. 81-86, 1998.
9. Lee, S. and Crawford, M.M., "Unsupervised Classification for Multi-sensor Data in Remote Sensing using Markov Random Field and Maximum Entropy Method", *Proceedings of the 1999 International Geoscience and Remote Sensing Symposium*, Hamburg, Germany, pp. 1200-1202, 1999
10. Brandt TSO and Paul M. Mather, "Classification of Remotely Sensed Imagery Using Markov Random Fields", Conference Proceedings of ACRS 2000, <http://www.gisdevelopment.net/aars/acrs/2000/ts12/ts12001.shtml>
11. Masalmah, Y.M. "Statistical Modeling of Clutter in Hyperspectral Data using 3D Markov Random Fields", [http://mayaweb.upr.clu.edu/crc/crc2002/papers/Masalmah\\_Yahya1.pdf](http://mayaweb.upr.clu.edu/crc/crc2002/papers/Masalmah_Yahya1.pdf)
12. Descombes, X., R. D. Morris, J. Zerubia and M. Berthod "Estimation of Markov Random Field Prior Parameters Using Markov Chain Monte Carlo Maximum Likelihood", *IEEE Trans on Image Processing*, vol. 8, No. 7, JULY, pp954-963, 1999
13. Çesmeli, E. and D.L. Wang, "Texture Segmentation Using Gaussian-Markov Random Fields and Neural Oscillator Networks" *IEEE Trans on Neural Networks*, vol. 12, No. 2, March 2001, pp 394-404
14. Hearst, M. "K-Means Clustering", UCB SIMS, Fall 1998; <http://www.sims.berkeley.edu/courses/is296a-3/f98/>
15. Li, S.Z. "Markov Random Field Modeling in Computer Vision" ISBN 0-387-70145-1 Springer-Verlag 1995.
16. Zoltan, K., B. Marc and Z. Josiane "A Hierarchical Markov Random Field Model and Multi-temperature Annealing for Parallel Image Classification" <ftp://ftp.inria.fr/INRIA/publication/publi-pdf/RR/RR-1938.pdf>
17. Bader, D.A., J. Jálá and R. Chellappa "Scalable Data Parallel Algorithms for Texture Synthesis and Compression using Gibbs Random Fields", Department of Electrical Engineering, University of Maryland, Oct 4, 1993 <http://www.umiacs.umd.edu/research/EXPAR/papers/3123/3123-html.html>
18. Chen, C.H. and G. G. Lee "On Digital Mammogram Segmentation and Microcalcification Detection Using Multiresolution Wavelet Analysis", *Graphical Models and Image Processing*, vol. 59, No. 5, Sep, pp. 349-364, 1997
19. Kaymak, U. "A Unified Approach for Practical Applications of Fuzzy Clustering" [www.few.eur.nl/few/people/kaymak/publications/bnaic2000a\\_kaymak.pdf](http://www.few.eur.nl/few/people/kaymak/publications/bnaic2000a_kaymak.pdf)



This page is intentionally left blank

# **Automatic Target Segmentation**

This page is intentionally left blank

# CHAPTER 15

## MULTISENSOR AUTOMATIC TARGET SEGMENTATION

Mark R. Stevens, Magnús Snorrason

*Charles River Analytics, 625 Mount Auburn St., Cambridge, MA 02138*

SENGVIENG AMPHAY

*AFRL/MNGI, 101 W. Eglin Blvd., Eglin AFB, FL 32542*

Fusing information from sensors with very different phenomenology is an attractive and challenging task for automatic target acquisition (ATA) systems. Sensor fusion improves results when correct target detections correlate between sensors while false alarms do not (due to different properties of targets such as shape and signature of targets). In this paper, we present a series of algorithms for detecting and segmenting targets from their background in passive millimeter wave (PMMW) and laser radar (LADAR) data. PMMW sensors provide a consistent signature for metallic targets, however their angular resolution is too limited to support further target classification. LADAR sensors provide the ATA systems with high angular resolution and 3-dimensional geometric shape information supporting accurate target identification. However, the shape-based segmentation can give very high probability of false alarm under structured clutter scenarios. Sensor fusion techniques are applied with the goal of maintaining high probability of detection while decreasing the false alarm rate.

### 1 Introduction

Sensor fusion is an attractive option for reducing false alarm rates while retaining high probability of detection. This statement leads to the problem of how to best fuse the information. A wide range of sensor fusion methods exist [1, 2, 3] and have been successfully applied to automatic target detection and segmentation. However, the best fusion techniques for a given domain, a set of sensors, a dataset, and a set of algorithms for the specific sensors are generally not immediately obvious.

The focus of this paper is an empirical performance comparison of fusion algorithms for combining the output of segmentation algorithms for PMMW, LADAR reflectance, and LADAR range imagery. First, the performance for each image type is evaluated independently on a substantial dataset containing co-registered images of the same target arrays. Both LADAR segmentation algorithms are capable of reliably locating almost every target instance in the dataset, but at a very high false alarm rate. The PMMW algorithm produces two to seven times fewer false alarms while maintaining a high detection rate, but that false alarm rate is still too high. Therefore, fusion across every combination of segmentation algorithm output was tried (PMMW & LADAR range, PMMW & LADAR reflectance, LADAR range & reflectance, and all three). Four different pixel-level fusion methods were applied in each case: 1) binary “AND”, 2) Bayesian fusion, 3) Dempster-Shafer, and 4) majority voting. We then compare and contrast the different combination mechanisms. The results indicate that sensor fusion is critical to robust performance in the presence of structured clutter, and a variety of fusion algorithms should be considered to find the right method for a given dataset.

### 2 Background

The most relevant previous work fuses PMMW and LADAR range for target detection [4]. In this work, jump edges in the LADAR range (discontinuities in range in the vertical and horizontal direction) were measured and fused with temperature gradients in

the PMMW imagery. Dempster-Shafer evidence combination was used to reduce the number of false alarms and increase the number of correct detections. On the general topic of sensor fusion, [1] summarizes past work and notes that typically sensor fusion has emphasized single modality sensors, with comparatively little work on different sensor modalities. The paper goes on to state that relating data from different modalities is more difficult, in part because of issues of sensor alignment and registration. While [5] and [6] have examples of successful mixed-modality fusion, this is still a young research area. Dasarathy has rigorously studied fusion at the feature and decision level in order to replace multiple *ad hoc* methods with a formal methodology [7]. Waltz and Llinas' book on multi-sensor data fusion is still considered among the standards in the field [3].

Our previous work has also addressed multi-sensor target detection and recognition. We have used a model-based approach, where the pose of an object is iteratively recovered using heuristic search [8, 9, 10]. The error function being minimized measures the quality of match between 3D model-features and the projection of those features into color charge coupled device (CCD), forward looking infrared (FLIR) and LADAR imagery. Target recognition, sensor registration, and pose determination all happen simultaneously thus improving reliability and robustness. We have also performed pilot studies and refinements to all of the segmentation algorithms discussed in this paper [11, 12] as well as preliminary sensor fusion analysis (binary combination only) [11].

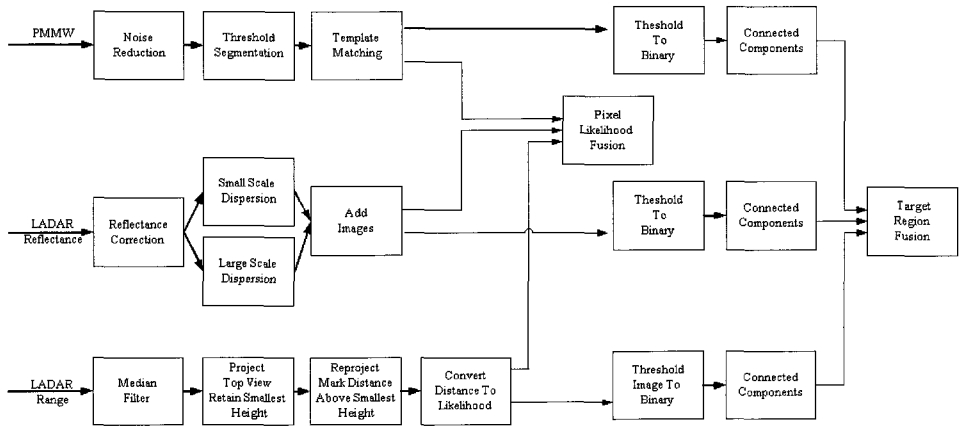
Finally, there exists an extensive body of literature covering the detection of targets using just LADAR range data or range and reflectance data. These approaches can be classified as feature based [13], model-based [14], and template based. Techniques for invariant analysis also have been examined [15].

### 3 The Segmentation Algorithms

The fundamental differences between an Automatic Target *Segmentation* (ATS) system and an Automatic Target *Detection* (ATD) system are in 1) the type of output produced, 2) the types of failures that can occur, and 3) the kind of evaluation that must be performed. An ATD system specifies *locations* of potential targets, typically in terms of image coordinates of the target centroid. The three types of failures that can occur are false negative mistakes (missed targets), false positive mistakes (false alarms), and location errors (target correctly detected but with inaccurate centroid). An ATS system specifies *regions* representing potential targets. The first two failure types are the same, missed targets and false alarms, but the third type is the *quality* of segmentation: does the segmented shape accurately match the image-truthed shape? To characterize segmentation accuracy, we use the quality evaluation criteria of over- and under-segmentation, as discussed in Section 3.4.

Another difference between ATS and ATD—at least in our work—is that ATS is just a module in a larger ATR system, while ATD is typically a stand-alone system. This leads to a difference in the amount of false positive mistakes that can be tolerated. In an ATR system, the ATS component provides potential targets to be run through a classifier that performs the desired level of target recognition. A substantial false positive, or *clutter* rate from an ATS system is therefore acceptable, so long as the following classifier can reliably label clutter (trees, bushes, non-military vehicles, etc.) as objects to be ignored from all further processing. A high *false alarm* rate from an ATD system is never acceptable, because it ruins the operator's confidence in the ATD systems (for human-in-the-loop systems) or causes erroneous decisions (in fully automated weapon systems).

Our ATS system uses three different segmentation algorithms to isolate probable targets in the sensor imagery (see Figure 1.) Each of the three algorithms processes a different sensor channel, and can segment the image at approximately the frame-rate of the sensor. Since the channels differ in the types of information they measure and the algorithms to treat this information are vastly different, we can treat the produced results of the three segmentation algorithms as independent. This is a required assumption for some of the fusion algorithms discussed in Section 4.



**Figure 1: Overview of segmentation algorithms**

Since the focus of this paper is not the segmentation algorithms, but rather the fusion of the segmentation results, only a short summary of each algorithm is provided (full details of the algorithms have been published elsewhere [11, 12]). All three algorithms consist of pixel by pixel classifiers, followed by thresholding and connected components grouping to produce target regions. Each algorithm produces two outputs: a segmentation mask designating regions of pixels belonging to the same potential target and a likelihood image representing the belief per pixel that it belongs to a target.

### 3.1 PMMW Segmentation

The PMMW imagery is segmented using a series of steps. Each step is more computationally demanding than its predecessor, but processes far fewer pixels. An analogy is a series of filters where each layer removes incorrect pixels and allows smaller amounts of information to pass through. The first algorithm in the pipeline is a noise reduction step. Analysis of the sensor data indicated that the lower order bits from the 12-bit input sensor image contained mostly noise and very little signal information. Hence, the four lowest order bits were removed from each pixel, converting the image to 8 bits. Next, a global image threshold (two standard deviations below the image mean) was used. Assuming the mode and mean are approximately equal (which we have verified on several different scenes), every pixel in the image is given one of two labels:

$$class(p) = \begin{cases} target & p < \mu - k\sigma \\ background & otherwise \end{cases}$$

where  $\mu$  is the mean of the entire image,  $p$  is a pixel in the image,  $\sigma$  is the standard deviation of the entire image, and  $k$  is a parameter dictating how far below the mean to place the threshold. Template matching is then used to detect potential target centroids by determining if a local region of image data closely matches the expected target signature. Finally, the correlation values produced by the template matcher are thresholded to binary and a connected components algorithm extracts target regions. Figure 2 shows an example PMMW image, the likelihood map output and the segmentation mask output. (Note that in this figure and the following ones, different gray-levels are shown in the segmentation masks to indicate each region's unique label.)

### 3.2 LADAR Reflectance Segmentation

The first step in LADAR reflectance segmentation is to pre-process the image to remove the sensor's fixed-pattern noise (the particular sensor used to collect data for this work suffered from vertical stripes due to gain variance across the laser diodes). Next, a multi-scale variance algorithm is used. The motivation for using variance is based on the observation that man-made objects—such as military vehicles—in LADAR reflectance images of outdoor scenes are characterized by one of three conditions:

- Reflectance values are significantly higher than the local image average
- Reflectance values are significantly lower than the local image average
- Reflectance characteristics are highly varying with a relatively high spatial frequency

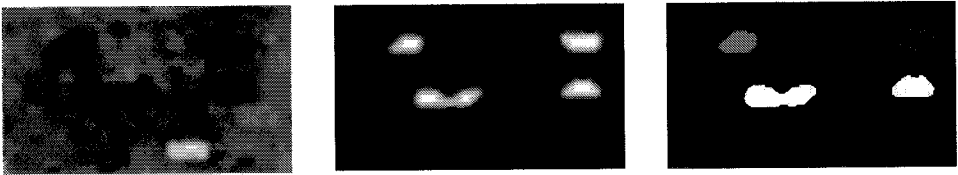
Multi-scale variance captures all three of these conditions and tends to segment target and some clutter regions from the rest of the image. Local variance estimates the deviation from the mean in a local neighborhood. Therefore, our three criteria above can be uniquely measured by computing local variance over the whole image using different window sizes. Computing variance using a small window (such as 3x3) emphasizes the high spatial frequency regions, while the variance measure for a large window (such as 11x11) emphasizes solid outlines of the regions that are significantly lighter or darker than average. One image is produced for each window size, and the images are added and thresholded to binary. The two window sizes and the threshold value were determined empirically, based on Receiver Operator Characteristic (ROC) curve analysis. The final step is to clean up the thresholded image (remove very small isolated regions and filling in small holes using binary morphology) and apply the connected components algorithm. Figure 3 shows an example LADAR reflectance image, the likelihood map and the segmentation mask output.

### 3.3 LADAR Range Segmentation

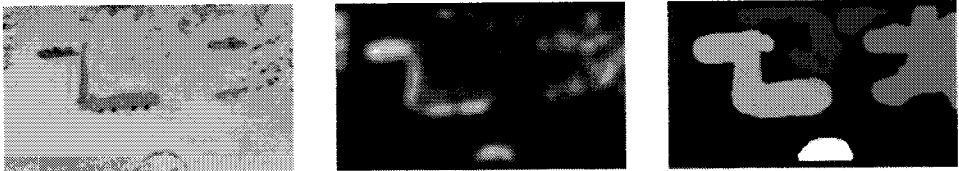
The first step in the LADAR range segmentation is to median filter the image to remove dropouts. Then the 3D location of each pixel in the scene is determined. The LADAR records range and angle in a spherical coordinate system, so each pixel is first converted to Cartesian coordinates  $\langle x, y, z \rangle$  using the depression & azimuth angles, placement of the pixel (row, column) within the focal plane array, range, and the angle resolution of the sensor. Once the coordinate transform has been performed, each range pixel has a corresponding  $\langle x, y, z \rangle$  component. We define  $x$  and  $y$  to be the horizontal ground plane and  $z$  to be the height above that plane.

Next, orthographic projection is used to generate a top-down view of the scene. When constructing this view, multiple range samples measured from the same vertical

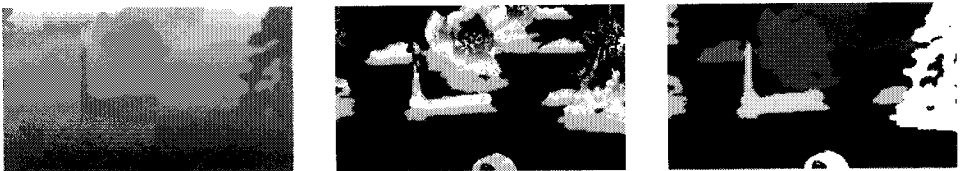
surface (such as the side of a target) are likely to project to the same pixel because the  $x$  and  $y$  values of these samples are tightly clustered. For such pixels, only the smallest height value is retained. This image therefore contains an estimate of the local ground plane around each such pixel. The range data is then projected into top-down view again, this time recording the distance of each pixel above the local ground plane. This distance is converted to a likelihood measure by assuming the target height is a Gaussian random variable having mean of 2.5 meters and a standard deviation of 3 meters. Finally, the likelihood is thresholded and converted to a target region mask using connected components. Figure 4 shows an example LADAR range image, the likelihood map and the segmentation mask output.



**Figure 2: Example PMMW segmentation showing input image (left), likelihood map output (center), and segmentation mask output (right)**



**Figure 3: Example LADAR reflectance segmentation showing input image (left), likelihood map output (center), and segmentation mask output (right)**



**Figure 4: Example LADAR range segmentation showing input image (left), likelihood map output (center), and segmentation mask output (right)**

### 3.4 Evaluating Segmentation Algorithms

We have decided upon five different performance metrics for our evaluation. These metrics were inspired by a significant and substantial range image segmentation evaluation performed on a large number of algorithms [17]. Table 1 defines the five different metrics.

Problems arose in the straightforward application of the metrics used in [17], due to symmetry. Originally, a correct detection was defined as  $T$  percent of a ground truth region matching a machine segmented region and  $T$  percent of that same machine segmented region matching the target [17]. For our ATS application, this is too stringent: we need to assure that all targets are labeled as targets, but are not concerned with assuring that the machine segmentation algorithm does not over explain the target. Therefore, we



reduced the symmetric relationship so that only the percentage of target overlap is measured.

Correct detections can be formally specified using set notation. If  $M$  is a region of pixels labeled as target by the ATS algorithm and  $N$  is a region labeled by hand as target, then the number of total correct detections is given as:

$$\text{Correct Detection} = \sum_{i,j} \begin{cases} 1 & |M_i \cup N_j| \geq T|M_i| \\ 0 & \text{otherwise} \end{cases}$$

which implies that the target region must overlap  $T$  percent of the segmented region. To prevent double counting, once a segmented region has been labeled as a correct detection, it is removed from further consideration as over and under segmented. For our experiments,  $T$  was set to 0.5 due to slight registration errors between the two sensors. An over segmentation is defined as:

$$\text{Over Segmentation} = \sum_i \begin{cases} 1 & \left| \bigcup_j M_i \cup N_j \right| \geq T|M_i| \\ 0 & \text{otherwise} \end{cases}$$

which implies that a target is over segmented if multiple segmentations overlap the true target region. An under segmentation is defined as:

$$\text{Under Segmentation} = \sum_j \begin{cases} 1 & \left| \bigcup_i M_i \cup N_j \right| \geq T|N_j| \\ 0 & \text{otherwise} \end{cases}$$

which measures over segmentation of image truthed regions.

True targets that do not fall into any of the above categories (correctly detected, over segmented, or under segmented) are then defined as missed targets. Finally, by symmetry, machine segmented regions that do not fall into any of the above categories are defined as clutter. Note that from the calculation of these five metric values, any possible evaluation metric can be computed.

Performance Measure	Description
Correct Detection	A machine-segmented region overlaps a target truth region such that at least 50% of the truth region is correctly labeled
Over Segmentation	The number of segmented regions matching to a single truth region
Under Segmentation	The number of truth regions matching to a single segmented region
Missed Targets	A truth target is not correctly detected, or is not over segmented, or is not under segmented
Clutter	A machine-segmented region is not correctly detected, or is not over segmented, or is not under segmented

**Table 1: Segmentation evaluation metrics**

## 4 The Sensor Fusion Algorithms

Numerous techniques exist for fusing information from multiple sensors [2]. Out of the plethora of techniques that exist, we have chosen to implement a subset of four and evaluate them using the metrics proposed in Section 3.4. These techniques, described in the following subsections, are: 1) binary *AND* fusion, 2) Bayesian fusion, 3) Dempster-Shafer fusion, and 4) majority voting fusion. Figure 5 shows example results for each of the four fusion methods on the same image pair.

### 4.1 Binary AND Fusion

Binary *AND* fusion takes as input the segmentation mask produced by each segmentation algorithm. Each region mask is then converted to a binary image by setting all target region pixels to 1 and all background pixels to 0. The resulting images are then all ANDed together to produce a new segmentation mask. A likelihood image is formed from this mask by passing a box filter over it and computing the average number of 1-bits in an 11x11 window. Connected components are then run on this likelihood image. Using binary *AND* fusion, we expect regions to be produced only where all sensors agree on the presence of a target. This should produce a very low clutter rate, but will miss a target if a single sensor misses that target.

### 4.2 Bayesian Fusion

Bayesian sensor fusion combines evidence according to Bayes chain rule with the assumption that the segmentation algorithms produce independent outputs. Once a likelihood image is produced, it is converted to a probability map by looking up the probability that each measurement is a target in a constructed histogram. This histogram is built using a series of hand-segmented images. This gives an estimate of  $P(t|m)$  where  $t$  is target and  $m$  is the value present in the histogram. Counting pixels in hand-segmented images is used to estimate of  $P(t)$ . Evidence is then combined across sensors using:

$$p(t | M_1 = m_{1k}, M_2 = m_{2k}, \dots, M_m = m_{mk}) = \frac{\prod_{k=1}^m p(t | m_k) p(m_k)}{\sum_{k=1}^m p(t | m_k) p(m_k)}$$

The output of this computation is a likelihood image. Connected components are then run on this likelihood image. Using Bayesian fusion, we expect only those regions with a high probability of target across all three channels (LADAR range, LADAR intensity and PMMW) to be retained in the result. However, if two segmentation algorithms have an extremely high likelihood and the third does not, the target region may still be preserved in the output. We also expect a higher clutter rate than binary *AND* fusion since clutter objects that correlate in two of the sensors are preserved.

### 4.3 Dempster-Shafer Fusion

Dempster-Shafer evidence combination [18] examines the belief in several hypotheses for each specific pixel. The segmentation algorithms all produce a likelihood image representing the belief that each pixel,  $P(T)$ , belongs to a target. We can estimate the likelihood the pixel is not part of a target by counting the number of neighboring pixels in the

connected components output. Pixels that do not have a large number of neighboring pixels classified as target have a high likelihood of being not target. We accomplish this by computing  $P(NT) = 1 - R / C$  for each pixel, where  $R$  is the number of target pixels in an  $11 \times 11$  region about each pixel and  $C$  is 121 (the total number of pixels). Finally, we measure the plausibility, or confidence that the pixel is explained properly, as  $P(\Phi) = 1 - P(T) - P(NT)$  where  $P(T)$  is the likelihood the pixel is target (from the likelihood segmentation image) and  $P(NT)$  is the likelihood the pixel is not target (from the neighbor analysis.) The three hypothesis (target, not target, and other) are then combined for each pair of segmentations according to the Dempster-Shafer evidence combination rule:

$$m_{ij}(A) = \frac{\sum_{X \cap Y = A} m_i(X) m_j(Y)}{1 - \sum_{X \cap Y = \phi} m_i(X) m_j(Y)}$$

where  $m$  represents a hypothesis about a pixel,  $X$  and  $Y$  are the measurements for each hypothesis, and  $A$  is the desired hypothesis output. This amounts to computing the total evidence for each of three hypotheses (target, not target, other) and normalizing by the inconsistent explanations (one sensor says target, the other not-target). This rule is applied to pairs of segmentation outputs. The result produces a new set of measurements that can then be combined with other segmentation results. For our specific three-hypothesis problem, the evidence combination rules are:

$$P_{ij}(T) = \frac{P_i(T)P_j(T) + P_i(T)P_j(\Phi) + P_i(\Phi)P_j(T)}{1 - P_i(T)P_j(NT) - P_i(NT)P_j(T)}$$

$$P_{ij}(NT) = \frac{P_i(NT)P_j(NT) + P_i(NT)P_j(\Phi) + P_i(\Phi)P_j(NT)}{1 - P_i(T)P_j(NT) - P_i(NT)P_j(T)}$$

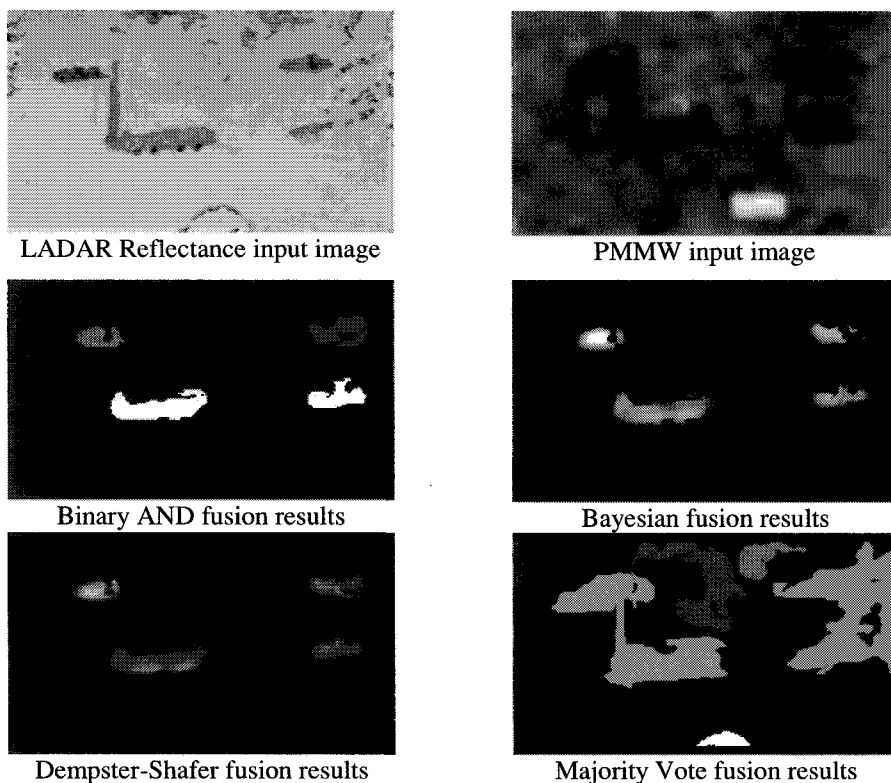
$$P_{ij}(\Phi) = \frac{P_i(\Phi)P_j(\Phi)}{1 - P_i(T)P_j(NT) - P_i(NT)P_j(T)}$$

Note that these values always sum to one. The best explanation for the pixel is the hypothesis with the highest belief. Only the pixels labeled as target are retained and placed in a likelihood image and connected components can then be used to produce a set of regions. Using Dempster-Shafer fusion, we expect to have a high probability of detection while minimizing clutter.

#### 4.4 Majority Voting Fusion

Majority voting takes as input each target region mask and converts it to binary in the same manner as described above for the binary AND fusion. A new image is produced that is the sum of all binary segmentation images. If each pixel's value is larger than half the total number of segmentation images used, the pixel is labeled as target. A likelihood map is formed from this image by passing a box filter over the binary result image and computing the average number of 1-bits in an  $11 \times 11$  window. Connected components are

then computed for this likelihood map. This algorithm will operate like a binary OR fusion if only two images are present. It is expected to have a high rate of detection, but also a high clutter rate.



**Figure 5: Sensor fusion results for one reflectance & PMMW image pair**

## 5 Results

We have applied the three segmentation algorithms and four fusion algorithms to a database containing 519 target instances of 9 different target classes (including clutter) in 95 images. The dataset consists of numerous static scenes as well as scenes where multiple vehicles are moving. While the optical axes of the LADAR and PMMW sensors were not perfectly boresight-aligned, some effort was made to have the axes coincide. Slight errors in registration were corrected by hand selecting tie-points in all pairs of images and computing a warping of one image onto another. This allows pixel level sensor fusion to take place. The images were all hand segmented into target vs. non-target regions to provide image truth information.

First, each segmentation algorithm was applied to each image and then evaluated using the segmentation evaluation criteria discussed in Section 3.4, producing the results in Table 2. The detection rate is the number of correct detections over the total number of targets present. The clutter rate is the number of clutter objects over the total number of regions detected. Note that while very few targets are missed, the clutter count for each sensor is high and there are many over and under segmentations (meaning multiple targets

are grouped into the same region or multiple regions cover the same target). Since the over and under segmented regions further characterize the correct detections, the correct detections plus clutter is the total number of regions extracted by the algorithm.

The four different sensor fusion techniques are then applied to all combinations of segmentation algorithm outputs (4 combinations) giving 16 different fused results. Table 3 shows all of the fusion results, and Figure 6 shows the detection rate versus clutter rate.

Using the definitions for “correct detection” and “clutter” from Table 2, we define the detection rate as:

$$\text{Detection Rate} = \frac{\text{Correct Detections}}{\text{Correct Detections} + \text{Missed Targets}}$$

and the clutter rate as:

$$\text{Clutter Rate} = \frac{\text{Clutter}}{\text{Correct Detections} + \text{Missed Targets} + \text{Clutter}}$$

There are several interesting trends in this table which are highlighted in the accompanying figure:

- LADAR range by itself produces a near perfect detection rate (0.998), but with a very high clutter rate (0.76). For any fusion method other than voting, adding PMMW information lowers the clutter rate by about 0.65, while only lowering the detection rate by about 0.05. The further addition of LADAR reflectance lowers both detection and clutter rates by about 0.01.
- In all cases, majority vote fusion produces the highest probability of detection. However, using this type of fusion also produces a high clutter rate. (Result A in Figure 6.)
- The second highest detection rate is always produced by the Bayesian fusion. For PMMW and range (with or without reflectance), this coincides with low clutter rate. (Results B in Figure 6 represents the overall best jointly optimized detection and clutter rates.)
- The least useful combination for decreasing clutter rate is to fuse LADAR reflectance and LADAR range segmentation (Result C in Figure 6). This result makes intuitive sense: the information for range and reflectance comes from the same LADAR sensor and is therefore more correlated with each other than with information from the PMMW sensor.

Sensor	Detection	Clutter	Correct	Over	Under	Missed	Clutter
	Rate	Rate	Detection	Segment	Segment	Targets	
PMMW	0.96	0.35	469	12	12	21	248
Reflectance	0.99	0.45	516	39	39	3	415
Range	1.00	0.76	518	88	88	1	1677

**Table 2: Segmentation results of each algorithm without fusion**

Sensors	Fusion	Detection Rate	Clutter Rate	Correct Detections	Over Segment	Under Segment	Missed Targets	Clutter
PMMW & Reflectance	Binary AND	0.94	0.14	451	23	11	31	71
	Bayesian	0.97	0.36	469	14	15	13	261
	DS	0.93	0.15	450	13	12	32	79
	Voting	0.99	0.68	479	44	17	3	1012
PMMW & Range	Binary AND	0.95	0.09	459	16	11	23	48
	Bayesian	0.97	0.12	467	25	12	15	66
	DS	0.94	0.11	453	11	13	29	57
	Voting	0.99	0.56	479	34	13	3	617
Reflectance & Range	Binary AND	0.99	0.58	513	44	36	6	696
	Bayesian	1.00	0.57	517	25	31	2	678
	DS	0.93	0.63	485	21	32	34	820
	Voting	1.00	0.38	519	2	73	0	314
PMMW & Reflectance & Range	Binary AND	0.93	0.08	448	23	11	34	39
	Bayesian	0.96	0.10	464	15	12	18	52
	DS	0.94	0.11	452	12	12	30	54
	Voting	0.99	0.43	478	29	13	4	358

Table 3: Segmentation fusion results, with clutter rates below 0.15 highlighted

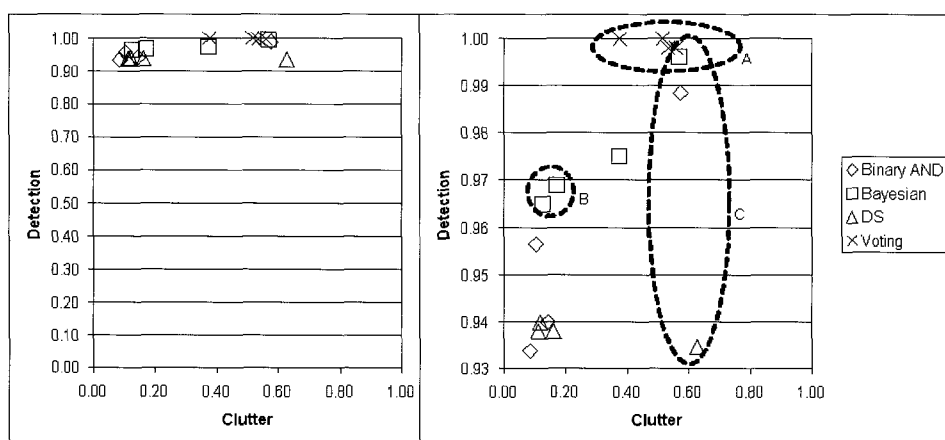


Figure 6: Probability of detection versus probability of clutter for all four fusion algorithms, same data shown with full scale Y-axis (left) and zoomed-in Y-axis (right).

## 6 Conclusions

We have presented a series of segmentation algorithms for a multisensor ATS task. In our results, LADAR range by itself produced a near perfect detection rate, but with a very high clutter rate. Adding information from a PMMW sensor lowered the clutter rate significantly, while only slightly lowering the detection rate. The further addition of LADAR reflectance information had little effect, slightly lowering both detection and clutter rates. Four fusion methods were tried, with Bayesian fusion performing best in terms of jointly optimizing detection and clutter rates. Dempster-Shafer fusion and binary AND fusion also performed well, while majority voting fusion did not.

These results also demonstrate the need for future work. Specifically, object level fusion should be examined to measure attributes of segmented regions in an effort to further reduce clutter and avoid the inherent problems associated with pixel level sensor fusion

## 7 Acknowledgments

Air Force Research Labs (AFRL) sponsored this work under SBIR contract number F08630-00-C-0027.

## 8 References

1. Aggarwal, J.K., "Multisensor Fusion for Automatic Scene Interpretation," Analysis and Interpretation for Range Images, R.C. Jain and A.K. Jain, Editors. 1990, Springer-Verlag.
2. Hall, D.L., "Mathematical Techniques in Multisensor Data Fusion," Artech House 1992, Norwood, MA.
3. Waltz, E. and J. Llinas, "Multisensor Data Fusion," Artech House, Inc.1990, Norwood, MA.
4. Lin, C., S. Amphay, and B. Sundstrom, "Sensor fusion with passive millimeter wave and laser radar for target detection," SPIE AeroSense 1999. Orlando, FL.
5. Magee, M.J., et al., "Experiments in Intensity Guided Range Sensing Recognition of Three-Dimensional Objects," IEEE Transactions on Pattern Analysis and Machine Intelligence, 1985. 7(6): p. 629-637.
6. Stentz, A. and Y. Goto, "The CMU Navigational Architecture," ARPA: Image Understanding Workshop 1987. Los Angeles.
7. Dasarathy, B.V., "A trainable decisions-in decisions-out (DEI-DEO) fusion system," SPIE AeroSense. 1998. Orlando, FL.
8. Stevens, M.R. and J.R. Beveridge, "Interleaving 3D Model Feature Prediction and Matching to Support Multi-Sensor Object Recognition," International Conference on Pattern Recognition. 1996. Austria.
9. Stevens, M.R. and J.R. Beveridge, "Precise Matching of 3D Target Models to Multisensor Data," IEEE Transactions on Image Processing, 1997. 6(1): p. 126-142.
10. Stevens, M.R. and J.R. Beveridge, "Multisensor Occlusion Reasoning," 14th International Conference on Pattern Recognition. 1998. Brisbane, Australia.
11. Stevens, M., et al., "ATA algorithm suite for co-boresighted PMMW and LADAR imagery," SPIE AeroSense 2001.
12. Stevens, M. and M. Snorrason, "Signature Analysis of Targets in PMMW Imagery," IEEE Workshop on Computer Vision Beyond the Visible Spectrum, 2001.
13. Caglayan, A.K., M. Snorrason, and H. Ruda, "Multi-Level Autonomous Target Identification," SBIR Phase II Final report for WL/MNGA,WL-TR-96-7029, April 1996.

14. Wellfare, M. and K. Norris-Zachery, "Characterization of Articulated Vehicles using LADAR Seekers," SPIE AeroSense 1997, Orlando FL.
15. Arnold, G.D. and I. Weiss, " Detection and recognition in LADAR with invariants and covariants," SPIE AeroSense, 2001.
16. Jain, A.K., "Fundamentals of Digital Image Processing," Prentice-Hall 1989, Englewood Cliffs, NJ.
17. Hoover, A., et al., "An Experimental Comparison of Range Image Segmentation Algorithms," T-PAMI, 1996. 18: p. 673-689.
18. Shafer, G. and R. Logan, "Implementing Dempster's Rule for Hierarchical Evidence," Artificial Intelligence, 1987. 33: p. 271-298.



This page is intentionally left blank

# **Neural Networks**

This page is intentionally left blank

## CHAPTER 16

### CATEGORY CLASSIFICATION USING NEURAL NETWORKS

Michifumi Yoshioka, Toru Fujinaka, and Sigeru Omatu

*Osaka Prefecture University  
599-8531, Sakai, Osaka, Japan*

In this chapter, we deal with the land cover mapping using neural networks and introduce some methods to improve the classification accuracy. These methods are competitive neural networks, Jeffries-Matusita distance, and textures. Competitive neural networks introduced here are Learning Vector Quantization and Self Organizing Map which are powerful tools for categorization. Jeffries-Matusita distance and textures are also adopted for the feature selection or extraction from pixels of the target image which are an important factor for the classification accuracy. Finally, we will show some simulation results to confirm the effectiveness of these methods.

#### 1. Introduction

Remote sensing data includes reflectional and radiational characteristics of naturally occurring features found on earth. It is possible to categorize the data by analyzing its spectral information. The categorization employs a pattern recognition method, where a category corresponds to a pattern. There are two types of recognition methods, i.e., supervised and unsupervised methods. The former method picks up some training data from the remote sensing data in advance and defines a category for each of them. This method then classifies the whole data based on a specific feature of each category. The latter method does not require selection of the training data, and this method classifies the remote sensing data based on similarities among the pixels<sup>1</sup>.

The maximum likelihood (ML) method has often been applied to the classification of remote sensing data as a supervised classification method<sup>2,3</sup>. The ML method uses a statistical approach based on the Bayesian theorem. It can be applied theoretically to any probability distribution. In real data processing, however, the discrimination function works under the assumption that the probability distribution for each category is an  $n$ -dimensional normal distribution where  $n$  denotes the number of spectral bands. The categories are determined from the land cover information. The probability distribution of real data for each category does not always follow normal distribution, which is essential to the conventional ML method.

Therefore, a non-parametric classification method becomes important for real data processing. Recently, a method using the multi-layered neural network has been applied to the analysis of remote sensing data<sup>4,5,6</sup>. The neural network is trained by the well-known error back-propagation (BP) method<sup>4,7</sup>. However, in the neural network approach to remote sensing analysis, the convergence of a learning process is sensitive to the selection of the training data set. Furthermore, the learning method often requires a large number of iterations and much computational time. Since the error back-propagation method is a black-box approach, it is difficult to give physical meaning to the weights connecting the neurons.

The remote sensing techniques have various advantages in monitoring the global environments on the earth surface in a short time. However, they also have a disadvantage that a successful monitoring depends on the weather condition. Rainy days are unfavorable in general, which becomes especially serious in the analysis of a tropical rain forest. Recently, multi-frequency and multi-polarimetric Synthetic Aperture Radar (SAR) data can be obtained by using airborne or spaceborne SAR sensor systems. Calibration techniques for the SAR data have been established, so that it may be used for such applications as terrain detection and environmental research. With several basic approaches, the SAR data classification methods have been gradually developed. These methods classify the SAR data into specific categories. The data observed by the quadruple polarization, i.e., the combination of horizontal and vertical polarization in the transmitting and receiving signals, includes the polarization backscattering coefficients and relative phases between polarizations<sup>8,9</sup>. Thus, better classification accuracy may be expected by using the polarimetric SAR data instead of mono-polarization data.

Supervised classification methods for the polarimetric SAR data can be divided into statistical and neural network approaches. The Bayesian methods reported in [10], [11], and [12] belong to the former approach. In recent years, the latter approach has been recognized for its classification performance. Algorithms using layered neural networks have been considered in [13], [14], and [15]. Algorithms using competitive neural networks have been considered in [2], [3], [16], and [17].

In general, the layered neural networks have a problem that they need much learning computation time in comparison with competitive neural networks. When adopting competitive neural networks to the classification of remote sensing data, it is difficult to accomplish sufficient classification accuracy using only basic learning algorithms. Hence, Hara et al.<sup>2</sup> proposed a classification method that incorporates both the neural network and the maximum likelihood methods. Ito and Omatu<sup>17</sup> use a classification method that integrates several competitive neural networks trained by a conscience algorithm. Although the classification accuracy is improved in these methods, they are more complex than the basic learning algorithms, and considerable adjustments are often needed in real applications. Furthermore, a feature vector must be selected appropriately before the classification phase. Otherwise, satisfactory accuracy cannot be obtained even if all the observation values are applied. In

Chen et al.<sup>13</sup>, experiments using airborne SAR (AIRSAR) data for several types of feature vectors have been carried out, and their accuracies have been compared. Unfortunately, there exists no detailed discussion up to date concerning a method of selecting a suitable combination.

In Section 2, we explain a classification method using a competitive neural network trained by only two Learning Vector Quantization (LVQ) algorithms<sup>18,19,20</sup>. We discuss a method in which a suitable feature vector is selected using the JM (Jeffries-Matusita) distance<sup>1</sup>. Moreover, we introduce a pseudo-relative phase between polarimetries in order to obtain higher classification accuracy. Finally, we propose a method using textures which are composed of pixels. Using textures, we can obtain more information in the same area, while the resolution for analysis decreases. The polarimetric SAR data observed by Shuttle Imaging Radar-C (SIR-C) is employed for the experiments<sup>21</sup>. The proposed and conventional methods are compared with respect to classification accuracy and quality of the resulting images. The neural network approach can be applied to more general area of image classification, while its effectiveness may depend on the specific features of the given data.

## 2. Polarimetric SAR Data Classification by Neural Networks

This section is concerned with the SAR data analysis in which the data are observed by measuring the active radar signals. Here, the data belong to the range of microwave, and it is difficult to get the geographical scenery like a picture. Therefore, a kind of data mining approaches should be used. In this section, we adopt various types of the LVQ methods to achieve this purpose. At last, we employ the method using texture to improve the classification accuracy.

### 2.1. The Classification Algorithm

We consider a classification algorithm using a competitive neural network. The competitive neural network is trained by the Learning Vector Quantization (LVQ) method<sup>18,19,20</sup>. The LVQ method will cyclically update the weight vectors so as to reward correct classifications and punish incorrect ones. Kohonen<sup>18,19,20</sup> proposed the LVQ1, LVQ2.1, and OLVQ1 (Optimized-learning-rate LVQ1) algorithms for the LVQ method. We will first describe the LVQ1 learning algorithm. Let  $w_i(t)$  denote a sequential value of  $w_i$  in the discrete-time,  $t = 0, 1, 2, \dots$ , where  $i = 1, \dots, M$ . After the training data  $x(t)$  is presented, competition occurs in the competitive layer where neuron  $c$  ultimately becomes the winner. Let  $x(t)$  belong to a category  $\omega_x$  and let neuron  $c$  be assigned to a category  $\omega_c$ . The LVQ1 updates the weight vectors as follows:

$$\left. \begin{aligned} w_c(t+1) &= w_c(t) + \alpha(t)[x(t) - w_c(t)] \\ w_i(t+1) &= w_i(t), \quad i \neq c \end{aligned} \right\} \text{if } \omega_x = \omega_c \quad (1)$$

$$\left. \begin{aligned} w_c(t+1) &= w_c(t) - \alpha(t)[x(t) - w_c(t)] \\ w_i(t+1) &= w_i(t), \quad i \neq c \end{aligned} \right\} \text{if } \omega_x \neq \omega_i \quad (2)$$

Here,  $\alpha(t)$  denotes the learning rate such that  $0 < \alpha(t) < 1$  and  $\alpha(t) \rightarrow 0$  if  $t \rightarrow \infty$ . This learning algorithm tends to push weight vectors away from the decision surfaces of Bayes rule<sup>22</sup>. To accelerate the convergence of the network, the OLVQ1 updates the weight vectors by means of the individual learning rate in every neuron. Learning rate  $\alpha_i(t)$  in a neuron  $i$  is updated as follows:

$$\left. \begin{aligned} \alpha_c(t+1) &= \frac{\alpha_c(t)}{1 + \alpha_c(t)} \\ \alpha_i(t+1) &= \alpha_i(t), \quad i \neq c \end{aligned} \right\} \text{if } \omega_x = \omega_i \quad (3)$$

$$\left. \begin{aligned} \alpha_c(t+1) &= \min \left( \frac{\alpha_c(t)}{1 + \alpha_c(t)}, \alpha_0 \right) \\ \alpha_i(t+1) &= \alpha_i(t), \quad i \neq c \end{aligned} \right\} \text{if } \omega_x \neq \omega_i \quad (4)$$

where the initial learning rate is given by  $\alpha_i(0) = \alpha_0$ ,  $i = 1, \dots, M$ . Thus, weight vectors in the center of the category will have rapidly decreasing learning rates and conversely those near category boundaries will have increasing rates. The LVQ2.1 approximates the Bayes rule by adjusting category boundaries. It updates only the weight vectors in a window region. The window region is defined by a parameter  $w$ ,  $0 \leq w \leq 1$ .  $E_u$  and  $E_v$  denote pattern distances from  $x(t)$  to the nearest weight vector  $w_u(t)$  and the second nearest weight vector  $w_v(t)$ , respectively. Either of the weight vectors,  $w_u(t)$  or  $w_v(t)$  belongs to the same category as  $x(t)$  and the other belongs to a separate category. It is known that the window region is represented as

$$\min \left( \frac{E_u}{E_v}, \frac{E_v}{E_u} \right) > \frac{1-w}{1+w} \quad (5)$$

Then the LVQ2.1 updates the weight vectors as follows:

$$\left. \begin{aligned} m_u(t+1) &= m_u(t) + \alpha(t)[x(t) - m_u(t)] \\ m_v(t+1) &= m_v(t) + \alpha(t)[x(t) - m_v(t)] \end{aligned} \right\} \quad (6)$$

In this chapter, we employ the following learning approach:

[Step 1] Move the weight vectors roughly by LVQ1 or OLVQ1.

[Step 2] Tune up the weight vectors in category boundaries by LVQ2.1.

In Step 2, an improvement in the classification ability is expected by re-training the weight vectors using LVQ2.1. The proposed classification method is applied to the SAR data with unknown category to the trained neural network, and the data is classified to the category where the winner neuron belongs.

## 2.2. Feature Vector Selection

Next, we describe the selection method of an appropriate feature vector using the JM distance for the classification. The JM distance is a divergence which measures the separability of a pair of probability distributions. The multi-looking polarimetric SAR data has a nine-dimensional real vector  $\mathbf{x}$  given by

$$\mathbf{x} = [S_{hv}S_{hv}^*, S_{vv}S_{vv}^*, S_{hh}S_{hh}^*, \Re\{S_{hh}S_{hv}^*\}, \Im\{S_{hh}S_{hv}^*\}, \Re\{S_{hh}S_{vv}^*\}, \Im\{S_{hh}S_{vv}^*\}, \Re\{S_{hv}S_{vv}^*\}, \Im\{S_{hv}S_{vv}^*\}]^T. \quad (7)$$

Here,  $S_{hv}$  denotes the complex scattering coefficient, where the transmitting and receiving polarization are vertical and receiving, respectively. The symbols  $\Re$  and  $\Im$  indicate the real and imaginary parts of a complex number, and  $*$  the complex conjugation. Note that  $\mathbf{x}$  is derived from the Stokes matrix. The relative phase  $\phi_{hhvv}^{(t)}$  between  $S_{hh}$  and  $S_{vv}$  is given by

$$\phi_{hhvv}^{(t)} = \tan^{-1} \frac{\Im\{S_{hh}S_{vv}^*\}}{\Re\{S_{hh}S_{vv}^*\}}, \quad 0^\circ \leq \phi_{hhvv}^{(t)} < 360^\circ \quad (8)$$

Since divergences using  $\phi_{hhvv}^{(t)}$  between categories may be small, we introduce pseudo-relative phase  $\phi_{hhvv}^{(c)}$  defined by

$$\phi_{hhvv}^{(c)} = \cos^{-1} \frac{\Re\{S_{hh}S_{vv}^*\}}{\sqrt{S_{hh}S_{hh}^*S_{vv}S_{vv}^*}}, \quad 0^\circ \leq \phi_{hhvv}^{(c)} < 360^\circ \quad (9)$$

Note that  $\phi_{hhvv}^{(c)}$  suffers from a mirror effect such that values of  $\phi_{hhvv}^{(t)}$  in  $180^\circ \leq \phi_{hhvv}^{(t)} < 360^\circ$  are reflected to the range of  $0^\circ \leq \phi_{hhvv}^{(c)} < 180^\circ$ . Relative phases  $\phi_{hhvv}^{(t)}$  and  $\phi_{hhvv}^{(c)}$  between  $S_{hh}$  and  $S_{vv}$ , and  $\phi_{hhvv}^{(t)}$  and  $\phi_{hhvv}^{(c)}$  between  $S_{hv}$  and  $S_{vv}$  are defined in the same manner<sup>16,17,23,24</sup>. An appropriate feature vector is selected from the nine elements in (7) and six relative phase elements according to the following method. First, we estimate one-dimensional probability density function (PDF) for each category in every element based on data collected from test sites. JM distances between all category pairs for each element are computed and their averages are obtained. Thus, the feature vector elements can be selected in order of magnitudes of the average JM distance.

## 2.3. Experimental Results

In this section we will show some numerical results for the preceding methods presented.

### 2.3.1. SAR Data and Study Area

The experimental study area is a rectangular stretch of land (approximately 8.7 km by 12.5 km) in Sakaide, Japan. Figure 1 illustrates the study area. The area



includes vegetation, industrial area, and densely populated districts. We perform experiments using this area in particular, as it is possible to obtain training and test data easily. "Data take 61:40" indicated in Table 1 was acquired in quadruple polarization mode (both horizontal and transmission and reception polarizations) at C-band (5.7 cm) and L-band (24.0 cm) wavelengths. The study area is covered by the SIR-C data. A gray scale image with  $696 \times 1,000$  pixels corresponding to the study area is shown in Fig. 2. In relatively high relief areas, foreshortening and layover effects may occur if side looking radar is used. Backscattering coefficients from these areas may be different from observation values estimated by target shapes and dielectric characteristics. Thus, in this chapter classification of only the smooth areas is considered. When referring to the topographical map corresponding to Fig. 1, the high relief areas in Fig. 2 are masked. These areas will not be classified but called masked areas. We define five categories ( $L=5$ ),  $\omega_1$  : factory,  $\omega_2$  : golf course,  $\omega_3$  : vegetation,  $\omega_4$  : urban, and  $\omega_5$  : water, respectively. Therefore, the number of output neurons  $M$  is a multiple of five in the competitive neural network. In the experiments, the training and test data set with 500 samples for each category was used. Figure 3 shows a location diagram of the training and test data where the white regions represent unselected areas. The training data are denoted by  $\blacksquare$ ,  $\blacktriangle$ , and  $\blacktriangledown$  where  $\blacksquare$  and  $\blacktriangle$  show 10 by 10 and 5 by 5 pixels areas, respectively and  $\blacktriangledown$  is a closed region of 25 pixels. The reason for using  $\blacktriangledown$  is that it shows an area of category  $\omega_2$  consisting of golf courses where the area shape is too complex to be described by squares. The training data consisting of 500 pixels for each category was obtained from a field survey. We adopt the entire part of  $\omega_1$  and  $\omega_2$  areas as the test data since there are not so many pixels. The remaining categories  $\omega_3$ ,  $\omega_4$ , and  $\omega_5$  range over wider areas whose scales are different. To make the adopted test data equal in size, we pick up one pixel at intervals of 3, 4, and 32 pixels from the  $\omega_3$ ,  $\omega_4$ , and  $\omega_5$  areas, respectively. Here, all classification methods employ the same training and test data in order to compare classification accuracies. The scoring matrix called confusion matrix<sup>2,12</sup> is computed for each classification result. We will evaluate the classification methods using average accuracy defined by averaging the trace values of the confusion matrix. It is not meaningful to evaluate confusion matrices using the overall accuracy and miss-classification rate in this experiment, since the test data for some parts of the study area are missing, as can be seen in Fig. 3.

Table 1. SIR-C data characteristics

Data take No.	61:40
Date	April 13, 1994
Time	4:26:00.981 (GMT)
Wavelength	C-band (5.7cm), L-band (24.0cm)
Polarizations	HH, VV, HV, VH
Scene center	$34^{\circ}9.4' N$ , $133^{\circ}58.6' E$
Incidence angle range	$26.3^{\circ} - 29.6^{\circ}$
Image size	Range 16.2 km/Azimuth 99.9 km
Digital image dimensions	1,296 pixels $\times$ 7,995 lines

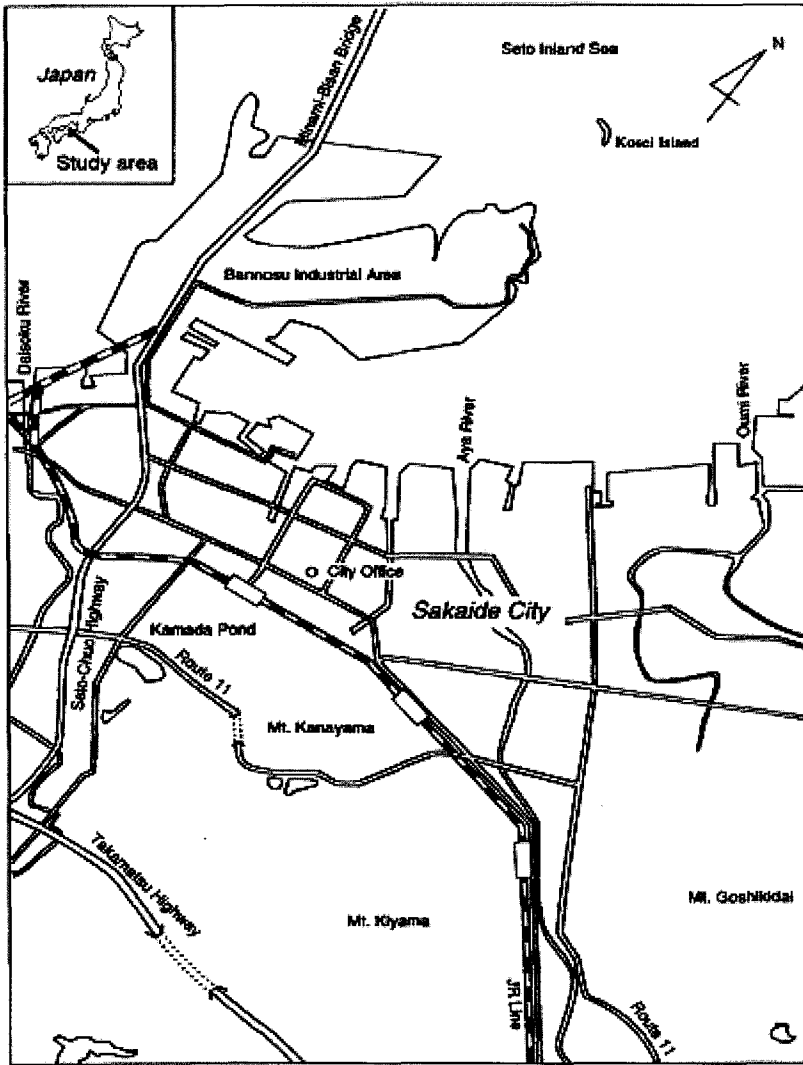


Fig. 1. Study area (Sakaide, Japan) for category classification.

### 2.3.2. Feature Vector

The JM distances for all category pairs are computed by using the test data of each element for appropriate selection of a feature vector. The average JM distances with all category pairs using C-band and L-band backscattering are calculated. The C-band and L-band backscattering data have a similar tendency. As the JM distances of the backscattering coefficients become larger, inclusion of these elements leads



Fig. 2. Gray scale image of SIR-C data for the study area.

to greater effectiveness. Moreover, the JM distance of the pseudo-relative phase  $\phi_{hhvv}^{(c)}$  is considerably larger than the other relative phases. Therefore, we select  $[10\log(S_{hv}S_{hv}^*), 10\log(S_{vv}S_{vv}^*), 10\log(S_{hv}S_{hv}^*), \phi_{hhvv}^{(c)}]$  for both C-band and L-band as the best combination, i.e., there are eight elements in the selected feature vector. The input layer also has eight neurons ( $N = 8$ ). Next, we note that  $\phi_{hhvv}^{(t)}$  is inferior to  $\phi_{hhvv}^{(c)}$  with respect to the divergence between categories. Investigation of the PDF of  $\phi_{hhvv}^{(t)}$  for each category using the C-band revealed that the phase  $\phi_{hhvv}^{(t)}$  for

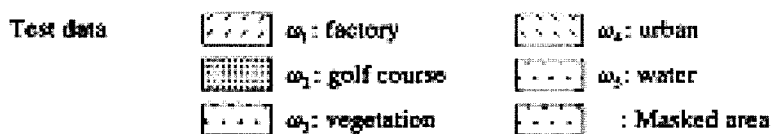
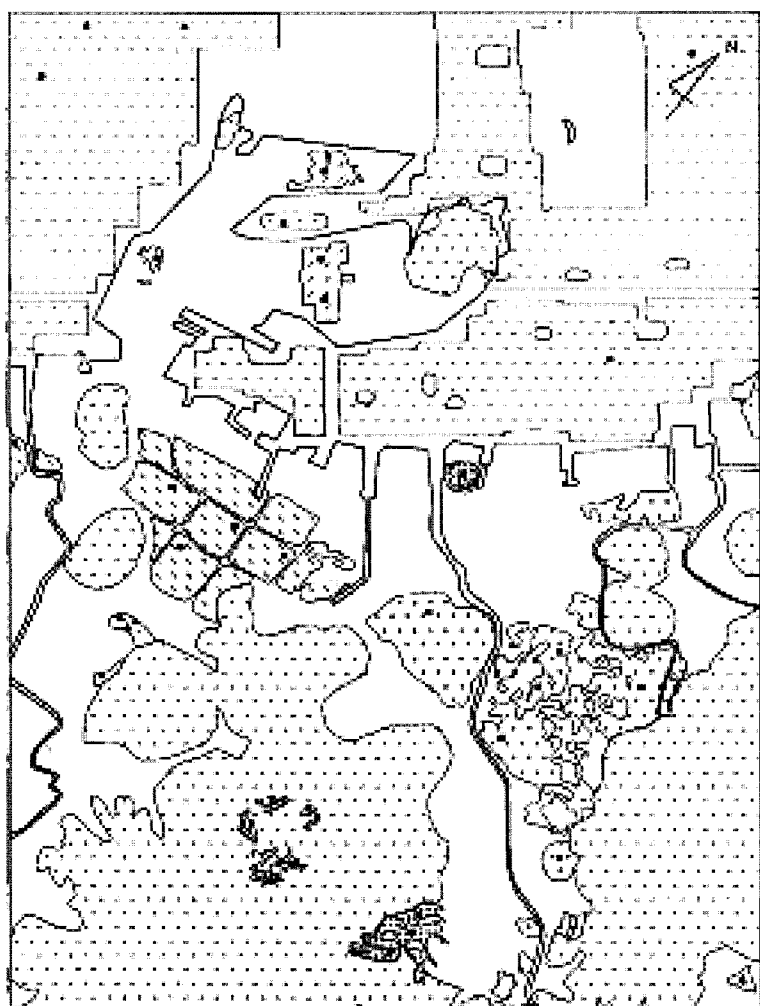


Fig. 3. Training and test data area

factory area ( $\omega_1$ ) differs from other categories by about  $180^\circ$ . The phase  $\phi_{hhvv}^{(t)}$  for water area ( $\omega_5$ ) has a concentration around  $0^\circ$ , and  $\phi_{hhvv}^{(t)}$  of  $\omega_2$ ,  $\omega_3$ , and  $\omega_4$  vary

with the PDF peaks. Hence, applying  $\phi_{hhvv}^{(t)}$  to a classification method would be significant. The differences of the PDFs for  $\phi_{hhvv}^{(c)}$  are significant among most of the categories compared with those of  $\phi_{hhvv}^{(t)}$ .

### 2.3.3. Classification Results

Table 2 indicates four kinds of learning algorithms applied to the LVQ methods in the experiments. Each algorithm includes either LVQ1 or OLVQ1, and the latter two algorithms incorporate LVQ2.1 as the second step of the algorithm. The number of training iterations for each step of the algorithm is determined by preliminary experiments. We set the initial learning rates to 0.03 for LVQ1 and LVQ2.1 and to 0.3 for OLVQ1. The window parameter is set to 0.3 in LVQ2.1.

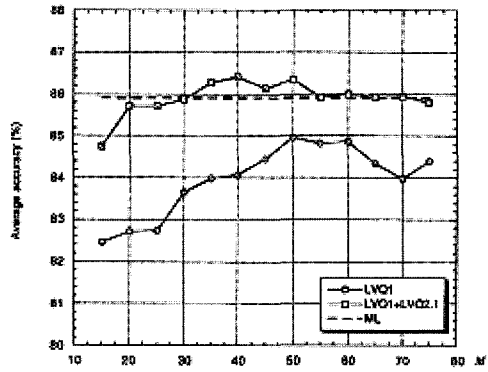
Table 2. Training iterations

Learning algorithms	Number of training iterations	
	Step1	Step 2
LVQ1	40,000	-
OLVQ1	10,000	-
LVQ1+LVQ2.1	40,000	40,000
OLVQ1+LVQ2.1	10,000	40,000

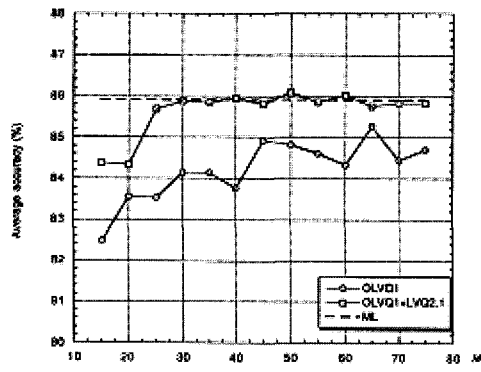
The LVQ methods are compared with the conventional methods of maximum likelihood (ML) method and the back-propagation (BP) method. The BP method employs three-layered neural network, where the number of neurons  $H$  in the hidden layer varies from 3 to 15. For the LVQ methods, five times as many, i.e., 15 to 75 neurons are used in the competitive layer for the LVQ methods. The selected eight-dimensional feature vectors are then adopted for each method and their test data classified. Figure 4 shows this comparison with respect to the average accuracies. Improvement of classification accuracy by adding LVQ2.1 to LVQ1 is 1.95% on average as shown in Fig. 4(a). Similarly, improvement of classification accuracy by adding LVQ2.1 to OLVQ1 is 1.42% on average as shown in Fig. 4(b). LVQ1+LVQ2.1 can produce higher average accuracies than the ML method. However, OLVQ1+LVQ2.1 can only produce similar average accuracies to the ML method. Thus, LVQ1+LVQ2.1 having  $M(35 \leq M \leq 70)$  produces better results than other methods as shown in Fig. 4(c). Especially, when the number of output neurons  $M$  is 40, it yields a highest accuracy reading of 86.40%. In most cases, the LVQ methods outperform the BP method. The resultant image of LVQ1+LVQ2.1 turns out to be more distinguishable than those obtained by the ML method where comparisons of the image quality are made.

### 2.4. The Method using Textures

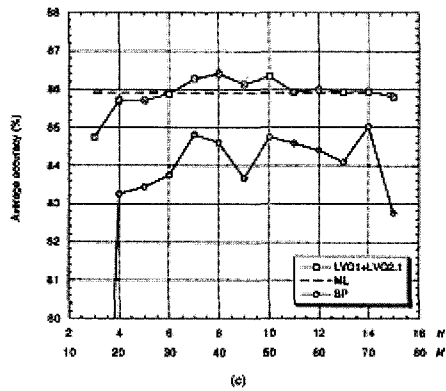
In this section, we will apply the method using textures to the same SIR-C data used in the previous section and show some simulation results.



(a)



(b)



(c)

Fig. 4. Comparison of classification accuracy among LVQ, ML and BP methods

## 2.4.1. Textures

In general, many feature indicators using textures are proposed. In this study, we adopt two types of indicators which have large variance in target categories. These texture indicators are known as Energy ( $E$ ) and Local homogeneity ( $L$ ) which are defined as follows:

$$E = \sum_{i,j} C_{ij}^2 \quad (10)$$

$$L = \sum_{i,j} \frac{1}{1 - (i - j)^2} C_{ij} \quad (11)$$

where  $C_{ij}$  denotes the joint probability of the appearance of pixel combinations which are composed of the gray levels  $i$  and  $j$  in the target area. In this study, we adopt 2 pixels as the distance between combined pixels and 8 directions as the direction from one pixel to another shown in Fig. 5.

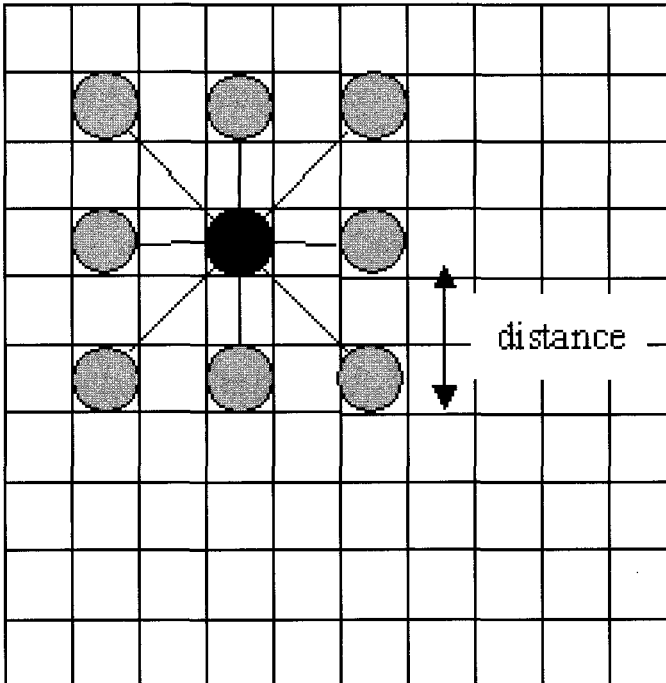


Fig. 5. Texture Structure.

### 2.4.2. Classification Algorithm using SOM

We adopt Self Organizing Map(SOM) as the classification method for the texture information because the distribution of texture features is too far from the normal distribution to apply statistical methods. SOM is composed of an input layer and a competitive layer figured in Fig. 6. In the input layer, the texture information made of pixels in the target area is fed and, in the competitive layer, the winner node is selected according to the Euclid distance between the input data and weights of competitive layer nodes. The node which has the weight of the minimum distance from the input becomes a winner. The weights of neighbour nodes surrounding the winner are changed according to the equs (12)-(14). After the training using all input data, some categories are automatically organized in the competitive layer.

$$U_j(t+1) = \begin{cases} U_j(t) + a(t)(U_j - E(t)) & j \in N(t) \\ U_j(t) & j \notin N(t) \end{cases} \quad (12)$$

$$a(t) = a_0 \left(1 - \frac{t}{T}\right) \quad (13)$$

$$N(t) = N_0 \left(1 - \frac{t}{T}\right) \quad (14)$$

where  $U(t)$ ,  $E(t)$ ,  $t$ ,  $T$ ,  $a(t)$ ,  $a_0$ ,  $N(t)$  and  $N_0$  denote a current weight of the  $j$ -th node in competitive layer, the current input data, the current iteration step, the total learning iteration, the current and initial learning parameter, the current and initial radius in the neighborhood area, respectively. The parameters used here are shown in Table 3.

Table 3. SOM Parameters

Number of neurons in the competitive layer	7×7
Total learning iteration	56000
Initial learning parameter	0.3
Initial radius of neighborhoods	3

### 2.4.3. Simulation Results

The results of simulations are shown in Table 4 and an example of classification result is shown in Fig. 7 and Table 5. In this case, we try five simulations and adopt the average of these simulations as the final result since the weights used in SOM are randomly initialized. The simulation result shows that the classification accuracy is improved by about 10% comparing with the methods mentioned in the previous sections by introducing textures while the resolution decreases.



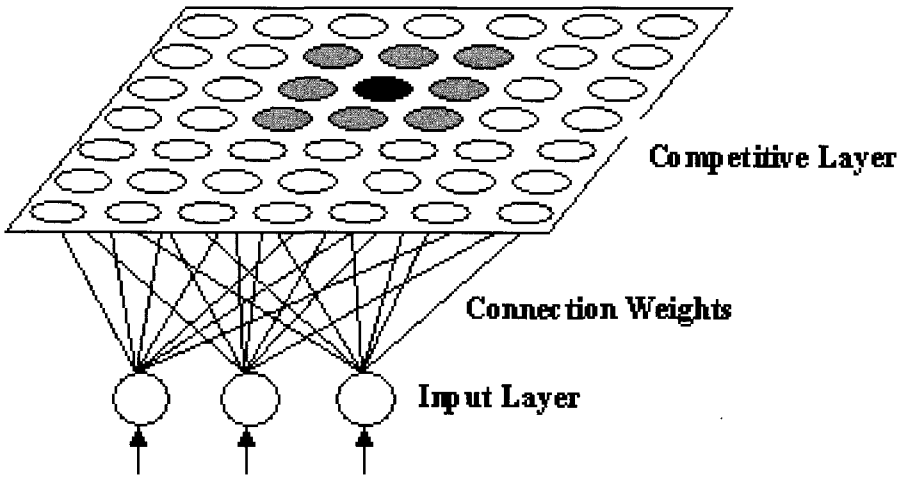


Fig. 6. SOM Structure.

Table 4. Classification Accuracy using Texture

Simulation No.	Accuracy(%)
1	97.4
2	95.0
3	96.6
4	96.9
5	96.2
Average	96.4

Table 5. Classification Score (Number of regions classified to each category)

		Reference category			
		Water	Factory	Urban	Vegetaion
Classified Category	Water	1393	0	0	5
	Factory	0	23	14	0
	Urban	0	8	137	12
	Vegetaion	4	0	5	213
Total		1397	31	156	230

### 3. Conclusions

We have proposed a new classification method of the remote sensing data, which uses the self-organizing neural network constructed from several neural network blocks, corresponding to the categories. In the experimental study, these blocks worked well as vector quantizers. The unknown pixels are successfully classified as well. We could demonstrate that the proposed method generally yields higher classification accuracies compared to other methods. If all neurons in the competitive layer have equal firing probability, the network works well as the vector quantiza-

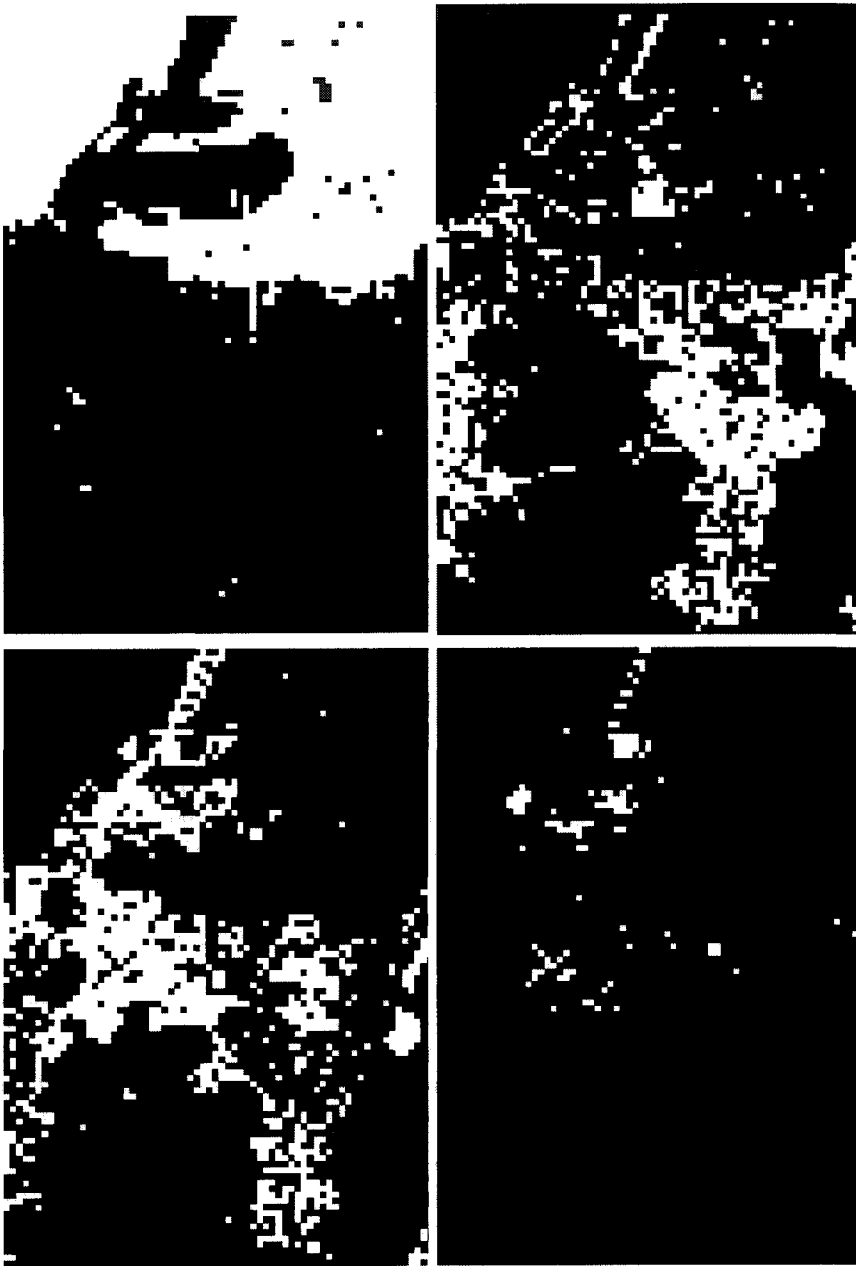


Fig. 7. Simulation result (Upper Left: Water, Upper Right: vegetation and golf course, Lower Left: Urban, Lower Right: factory)

tion of the training data. Then we proposed a classification method that employs a competitive neural network and a procedure to select a suitable feature vector.

After a preliminary experiment, three backscattering coefficients and a pseudo-relative phase between  $S_{hh}$  and  $S_{vv}$  are selected, using the JM distances. Among the methods considered, the neural network trained by LVQ1+LVQ2.1 produces the best classification results. Finally, we applied the method using textures to the SAR-C data and the improvement of the classification accuracy was confirmed by the simulation results. For future study, we will consider a classification algorithm which can classify the data into more detailed categories using multi-frequency and multi-polarimetric SAR data as well as physical backscattering models.

## References

1. Richards J. A. *Remote Sensing Digital Image Analysis, 2nd edn.* Springer-Verlag, Berlin, 246 (1993).
2. Hara Y., Atkins R. G., Yueh S. H., Shin R. T. and Kong J. A., *Application of Neural Networks to Radar Image Classification.* *IEEE Transactions on Geoscience and Remote Sensing* **32**, 100 (1994).
3. Hara Y., Atkins R. G., Shin R. T., Kong J. A., Yueh S. H. and Kwok R., *Application of Neural Networks for Sea Ice Classification in Polarimetric SAR Images.* *IEEE Transactions on Geoscience and Remote Sensing* **33**, 740 (1995).
4. Heermann P.D. and Khazenie N., *Classification of Multispectral Remote Sensing Data Using a Back-Propagation Neural Network.* *IEEE Transactions on Geoscience and Remote Sensing* **30**, 81 (1992).
5. Yoshida T., Omatu S. and Teranishi M., *Pattern Classification for Remote Sensing Data Using Neural Network.* *Transactions of the Institute of Systems, Control and Information Engineers* **4**, 11 (1991).
6. Yoshida T. and Omatu S., *Neural Network Approach to Land Cover Mapping.* *IEEE Transactions on Geoscience and Remote Sensing* **32**, 1103 (1994).
7. Hecht-Nielsen R., *Neurocomputing,* Addison-Wesley, New York, (1990).
8. Ulaby F. T. and Elachi C., *Radar Polarimetry for Geoscience Applications.* Artech House, Norwood, (1990).
9. Van Zyl J. J., Zebker H. A. and Elachi C., *Imaging Radar Polarization Signatures: Theory and Observation.* *Radio Science* **22**, 529(1987).
10. Lim H. H., Swartz A. A., Yueh H. A., Kong J. A., Shin R. T. and Van Zyl J. J., *Classification of Earth Terrain Using Polarimetric Synthetic Aperture Radar Images.* *Journal of Geophysical Research* **94**, 7049(1989).
11. Van Zyl J. J. and Burnette C. F., *Bayesian Classification of Polarimetric SAR Images Using Adaptive a Priori Probabilities.* *International Journal of Remote Sensing* **13**, 835 (1992).
12. Yueh H. A., Swartz A. A. and Kong J. A., *Bayes Classification of Terrain Cover Using Normalized Polarimetric Data.* *Journal of Geophysical Research* **93**, 15261 (1988).
13. Chen K. S., Huang W. P., Tsay D. H., and Amar F., *Classification of Multifrequency Polarimetric SAR Imagery Using a Dynamic Learning Neural Network.* *IEEE Transactions on Geoscience and Remote Sensing* **34**, 814 (1996).
14. Decatur, S. E., *Application of Neural Networks to Terrain Classification.* In: *Neural Networks, International Joint Conference at Washington D. C., U.S.A., June, IEEE, Piscataway*, 283 (1989).
15. Kanellopoulos I., Varfis A., Wilkinson G. G. and Megier J., *Land-cover Discrimination in SPOT HRV Imagery Using an Artificial Neural Network - a 20-Class Experiment.* *International Journal of Remote Sensing* **13**, 917 (1992).

16. Ito Y. and Omatu S., *Land Cover Mapping Method Using a Competitive Neural Network*. *Journal of the Japan Society of Photogrammetry and Remote Sensing* **35**, 5 (1996).
17. Ito Y. and Omatu S., *Category Classification Method Using a Self-organizing Neural Network*. *International Journal of Remote Sensing* **18**, 829 (1997).
18. Kohonen T., *Self-Organizing Maps*. Springer-Verlag, Berlin, 203 (1997).
19. Kohonen T., *Self-Organization and Associative Memory*. Springer-Verlag, New York, (1984).
20. Kohonen T., *An Introduction to Neural Computing*. *Neural Networks* **1**,3-16 (1988).
21. Stofan E. R., Evans D. L., Schmullius C., Holt B., Plaut J. J., Van Zyl J., Wall S. D. and Way J., *Overview of Results of Spaceborne Imaging Radar-C, X-band Synthetic Aperture Radar (SIR-C/X-SAR)*. *IEEE Transactions on Geoscience and Remote Sensing* **33**, 817 (1995).
22. Barndorff-Nielsen O. E., Jensen J. L., and Kendall W. S., *Networks and Chaos-Statistical and Probabilistic Aspects*, Chapman & Hall., London, 40 (1993).
23. Ito Y. and Omatu S., *Category Classification Using Self-organization Model*. In: *Noise and Clutter Rejection in Radars and Imaging Sensors, the 1994 International Symposium at Kawasaki, November, IEEE, Kawasaki, Japan*, 537 (1994).
24. Ito Y. and Omatu S., *Land Cover Mapping Method for Polarimetric SAR Data*. In: *Algorithms for Synthetic Aperture Radar Imagery IV, SPIE at Orlando, U.S.A., April, SPIE, Bellingham*, 388 (1997).

This page is intentionally left blank

## CHAPTER 17

# APPLICATION OF MULTIREOLUTION REMOTE SENSING IMAGE ANALYSIS AND NEURAL TECHNIQUES TO POWER LINES SURVEILLANCE

Elisabetta Binaghi, Ignazio Gallo\*, Monica Pepe\*

*Dep. of Information and Communication Science, University of Insubria, Varese -Italy.*

*\* National Research Council-ITC, Milan- Italy*

The larger availability of high resolution remotely sensed data, provided by novel aircraft and space sensors, offers new perspective to image processing techniques, but it introduces also the need for operational tools in order to completely exploit the potentialities of these data. These tools can be useful in many applicative contexts, in particular technological network surveillance, which involves specific requirements, such as accuracy in object recognition and positioning together with minimal demand of ground truth. The application presented deals with the recognition of features of interest for the surveillance of power transmission lines using IKONOS imagery. We proposed a methodology in which multi-scale and neural techniques are synergically combined to identify features at different scales and to fuse them for class discrimination. As seen in our experimental context, the results obtained on a pilot area in Northern Italy proved that the combination of multi-window feature extraction and neural soft classification produced a robust and flexible model that can act as a classifier of objects that vary in shape, size and structure.

### 1. Introduction

Recent advances in space techniques determine more and more diffused and diversified use of new types of data, in particular, very high resolution optical sensors such as Ikonos sensors and Earthwatch's Quick Bird<sup>1</sup>. The availability of these type of data offers new perspective to Remote Sensing creating the premise for new fields of application such as technological network surveillance and management.

The management and monitoring with high level of security of networks such as pipelines and power transmission lines include several tasks such as urban planning, water resources management, natural risk assessment. High resolution remote sensing imagery play a key role in all these activity.

In order to move towards adequate classification methodology one issue that has received particular attention within the remote sensing community is the development of contextual classification models in alternative to conventional quantitative classification techniques. The interest moves from the consideration that in high resolution imagery spectral values of neighboring pixels are strongly related or correlated since certain classes are likely to occur in the context of others and base their classification decisions on local and/or global spatial arrangement of pixel spectral values.

Several works have been proposed. Among early approaches, we found the use of local windows or surrounding pixels considered at a given distance and/or direction<sup>3</sup>. Other approaches based on relaxation labeling technique have been deeply investigated, dealing with context in a post-classification mode and addressing the critical problem of identifying land use patterns<sup>4</sup>. More recently works introduce the use of transforms methods, such as wavelet transform, for classifying urban patterns in high resolution remote sensing images<sup>5</sup>.

Despite the relevant achievements obtained, the use of contextual classification procedures is still limited by critical aspects, in particular complexity in methodology development, problem dependency, high demand of supervised data for training phase. In this work attention is focused on multi-scale approaches allowing to extract the spatial information from an image over a range of scales, from coarse details (large features), to fine details in local area. Multiscale image processing has a cognitive motivation: when we look at images with small and large objects, and/or low and high contrast objects, it can be advantageous to study them at several resolutions operating different zooming on the image; the appeal of this approach is that features that may go undetected at one resolution may be easy to spot at another<sup>6</sup>. Within the multiresolution approach the problem is addressed by two main strategies: - proposing methodology in which filters of fixed size are applied to iteratively scaled versions of the original image or -scaled versions of the original filter are applied to the original unvaried image<sup>7</sup>.

Various approaches have been reported in literature adopting different solutions for feature extraction and decision phase: image pyramids, filter banks and wavelet techniques have been deeply investigated achieving a significant interest in Remote Sensing<sup>8,9,10,11</sup>.

In this chapter we introduce contextual techniques and present a methodology for object recognition in high resolution remote sensing imagery. The salient aspect of the methodology is the integration of multi-scale and neural techniques to identify features at different scales and to fuse them for class discrimination. Results obtained by applying the methodology to the field of power transmission lines surveillance are reported. In order to satisfy application requirements, the solution proposed move from the objective of finding contextual methods of analysis capable of optimizing the balance between high accuracy and limited design and computational complexity and

providing also good generalization capabilities together with low training requirements in terms of examples provided and learning charge.

## 2. Contextual classification procedures

*Contextual classification procedures* can be categorized in a typology of context<sup>12</sup> by the use of raw as opposed to classified data, resulting in a major subdivision between contextual classifiers and contextual re-classifiers. The first branch are either based on the use of local windows<sup>13-15</sup> or consider the pixels at a given distance and/or direction<sup>16</sup>.

The second type of branch is based on the application of certain processes following the results of a preliminary per-pixel classification, such as in relaxation labeling and knowledge-based systems<sup>17-22</sup>: pixels are considered either separately (pixel-based), or arranged to form discrete objects (object-based) using connectivity, distance, direction and containment properties.

A common characteristic of these methods is the explicit representation of spatial relationships involving complex mathematical descriptions, whatever the representation framework may be.

An interesting experiment conducted by Paola and Showengerdt<sup>23</sup> consisted in providing spatial information within a classification procedure by presenting directly a  $3 \times 3$  window of image data as an input to a Multilayer Perceptron (MLP) classifier. Even within the limits of the experiment, which did not aim to define a complete contextual classification procedure, the important finding was that, although the training time required for each iteration was high, the network converged much faster due to the extra information used.

## 3. Classification Strategy

The classification strategy here proposed takes the origin from a previous work where a classification strategy based on MLP model and fixed windows was defined and tested<sup>24</sup>. It uses a moving window to evaluate the neighboring influences during classification in such a way that spatial relationships among the window pixels to be classified are not explicitly formalized. Instead, the corresponding window is directly presented as input to the neural network classifier. The classification strategy is area-based associating the input moving window, of fixed size and step, with a corresponding output window. Moreover the area-based character of the methodology, which provides the classification of a group of pixels rather than a single one, reduces complexity and computational costs, typical of contextual paradigms.

The results obtained in the experimental application were encouraging, however certain limitations of the strategy were identified in the use of a single window of fixed size to model the context.



This research project attempts to generalize the previously investigated approach in two perspectives:

- the use of a set of input concentric windows, instead of a single one, to evaluate the neighbouring influences during classification;
- to define the output of classification in terms of a target element consisting of one pixel or a group of pixels, providing for a strategy that would act either as a pixel-based or area-based contextual classification.

When dealing with windows both contextual classifiers and re-classifiers refer to a single pixel or to a group of pixels, i.e. they are pixel-based or area-based, and this choice affects the overall methodology configuration; our strategy generalises by introducing a more abstract concept, the *target element*, which represents the element to be classified defined as the fundamental  $K \times K$  frame (with  $K \geq 1$ ) around which contextual windows will be built and considered.

### 3.1 Contextual information extraction by cognitive – pyramid

Consistent with the cognitive foundation of contextual approaches, the question of determining the appropriate window size should be addressed both from the computational and human perceptual perspective. In remote sensing literature, works dealing with contextual per-pixel classifications calibrate the appropriate window size for operating on spectral data in a range from a 3 by 3 to a 9 by 9 matrix of pixels<sup>25,26,27</sup> considering that larger window sizes increase computational demands without increasing classification accuracy. However, cognitive-based studies suggest that humans require windows of considerably larger sizes than those used in digital image classification studies. This is confirmed in experiments conducted by Hodgson<sup>6</sup> showing that the photointerpreters use windows of about 40 by 40 pixels (corresponding to about 60 m by 60 m) for identifying urban patterns in high spatial resolution imagery.

To take into account both computational and cognitive perspectives our strategy adopts the solution of considering a large “perceptually-dimensioned” window centered on the target element, and then reducing size by compacting the neighboring information contained before entering the classifier.

We extend this base solution adding the consideration that human interpreters may simultaneously use dynamic windows of a variety of sizes to build evidence for recognizing image patterns<sup>6</sup> (Figure 1-A).

To include this perceptual aspect in the classification strategy the single window of fixed size is replaced by a set of concentric windows centred on the target element. Each concentric window is then resampled before entering the classifier (Figure 1-B).

The overall contextual information considered for each target element is therefore the result of the processes of enlarging the field of view (FOV) - starting from the target element - and compacting information at each step of the procedure.

The number of concentric windows and their sizes depend on the image resolution and the specific recognition target characteristics. For example, recognition of continuous targets requires larger context fields and more enlarging steps than those required for recognition of isolated, localized targets.

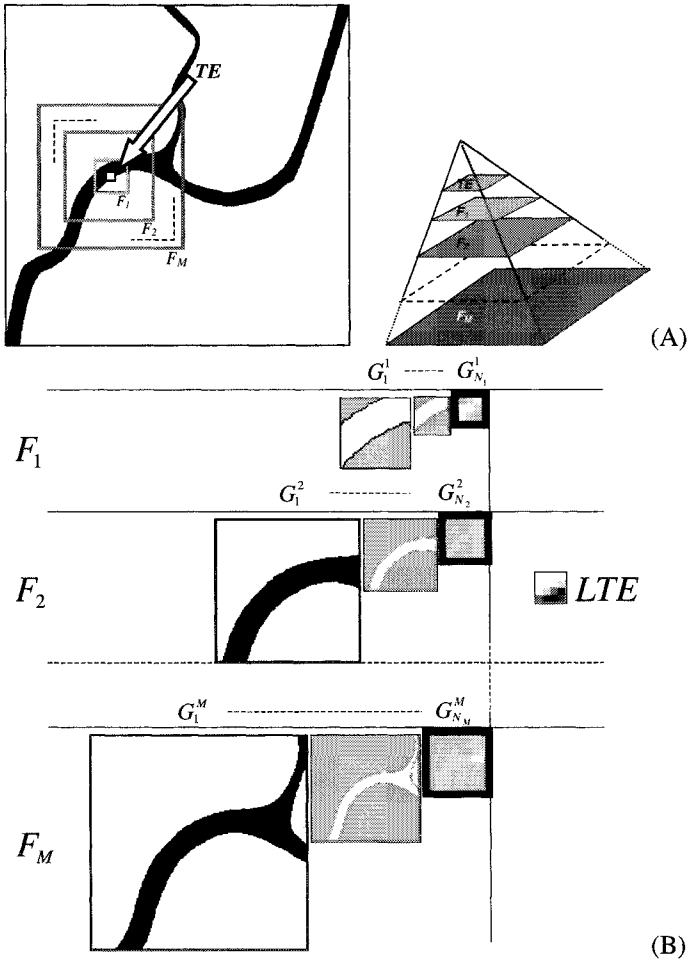
At each enlarging step ( $F_M$ ) the window dimension is independent of the former, so that the procedure can be extended to any object detection on the basis of perceptual needs alone.

When using a moving window to define the neighborhoods of a pixel, or a group of pixels, the problem of the border region needs to be addressed.

Although there are various possible solutions to this problem, we chose to reflect the source image pixels in the border area outward to effectively increase the size of the source image, taking the largest concentric window dimensions, thus providing a neighborhood for each target element, even in the border positions (see Figure 2). The rationale behind this choice is related to the use of cognitively-dimensioned windows, which are often considerably large, so any methodology that reduces the source image dimensions is unsuited to the purpose.

The compacting process is aimed at making the strategy computationally efficient, reducing the dimensions of window data that must simultaneously be considered for each step of the moving window procedure, besides reducing redundancy and noise among neighboring pixels by extracting some meaningful signals from the background. These objectives can be reached by using an image resampling pyramidal approach<sup>28</sup>.

In particular, the resampling technique chosen is the Gaussian image resolution pyramid<sup>29</sup>. It is a scale-space transform which makes it possible to efficiently include global, intermediate and local scales in the analysis without requiring a full convolution of the image, and avoiding the discontinuities characteristic of aliasing, due to resampling.



**Figure 1.** Contextual information extraction by cognitive – pyramid procedure.  
 (A) Around the *target element* (TE) to be classified a set of concentric windows  $\{F_1, \dots, F_M\}$  each of them with dimensions  $D_N^i, i = 1, \dots, M$ , is defined on the basis of the human interpreters' attitude during pattern recognition.  
 (B) Each concentric window  $F_i$  is then compacted by a Gaussian image resolution pyramid, starting from original image ( $G_1^i$ ) dimensions  $D_1^i \times D_1^i$  and reaching the root image ( $G_{N_M}^i$ ) dimensions  $D_{N_M}^i \times D_{N_M}^i$  before entering the classifier, to reduce data dimensions as well as redundancy and noise among neighbouring pixels

It is based on a multi-resolution image representation obtained through a recursive *reduction* - lowpass filtering and decimation - of the image data set.

Let  $G_0$  be a grey-level square image (see Figure 1-B), the Gaussian pyramid, for a decimation factor of 2, is defined according to Burt and Adelson<sup>8</sup> as:

$$G_k(i, j) = REDUCE(G_{k-1}) = \sum_{m=-2}^2 \sum_{n=-2}^2 w(m, n) \cdot G_{k-1}(2i+m, 2j+n)$$

where:

$k = 1, \dots, N$ ;  $N$  being the number of pyramid levels,

$G_N$  being the top, or *root*, of size  $u \times u$ ,

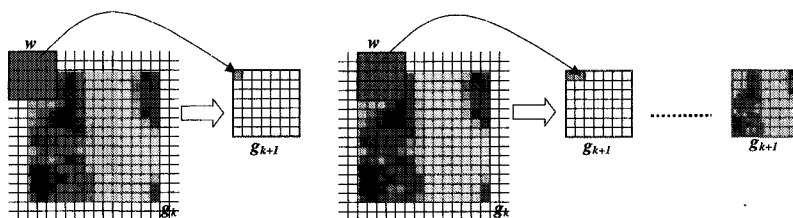
$i = j = 0, \dots, \frac{u \cdot 2^N}{2^k} - 1$  and

$$w(m, n) = \begin{bmatrix} 0.0025 & 0.0125 & 0.02 & 0.0125 & 0.0025 \\ 0.0125 & 0.0625 & 0.1 & 0.0625 & 0.0125 \\ 0.02 & 0.1 & 0.16 & 0.1 & 0.02 \\ 0.0125 & 0.0625 & 0.1 & 0.0625 & 0.0125 \\ 0.0025 & 0.0125 & 0.02 & 0.0125 & 0.0025 \end{bmatrix}$$

The number of columns and rows in the  $G_0$  image is  $D_0 \times D_0$ , related to the root dimensions by the following:

$$D_0 = u \cdot 2^N$$

The number of Gaussian-pyramid levels ( $N_i$ ;  $i = 1, \dots, M$ ) for each enlarging step ( $F_i$ ;  $i = 1, \dots, M$ ) is fixed with a view to the possible enhancing of the signal in relation to background, taking advantage of the reduction process, and stopping before any meaningful details are lost.



**Figure 2.** Solution adopted for the border region problem during the compacting process: the source image pixels in the border area were reflected outward to increase the size of the source image itself

### 3.2 Soft neural classification

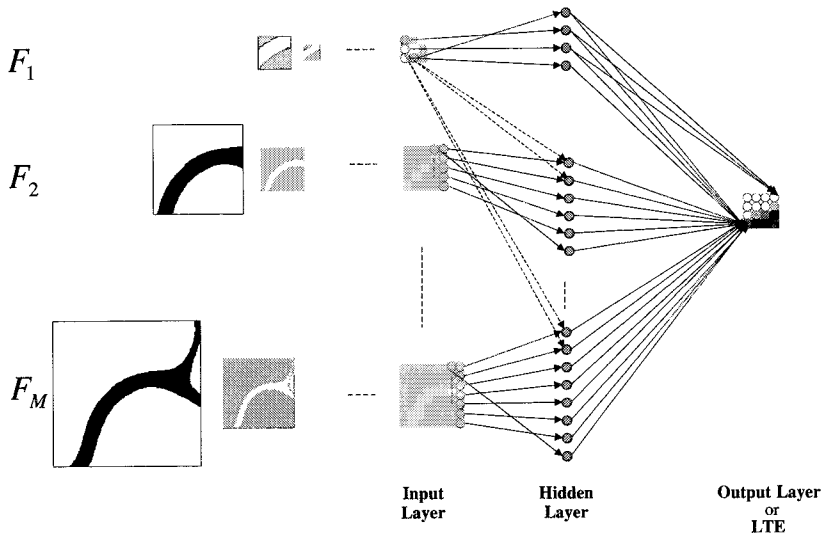
The application of the feature extraction strategy presented in the previous section provides inputs for the classification.

Our strategy adopts a soft classification paradigm, a choice motivated by the acknowledgment that the spatial variability of images results from the soft gradual character of the natural categories and that human interpretation is based on approximate modes of reasoning.

The strategy is based on the Multi-Layer Perceptron model: the use of this model is motivated by:

- the experimentally-proven effectiveness of the model when run as a soft classifier, dealing with intrinsic uncertainties in assigning elements to classes<sup>30,31</sup>;
- the well-documented capability of MLP in dealing with patterns described by complex features;
- the good performances demonstrated by the model in classifying window input data<sup>23</sup>;

The result of the context feature extraction procedure is constituted by the set of windows  $\{F_1, \dots, F_M\}$ , each of them with dimensions  $D_N^i, i=1, \dots, M$  (see Figure 1). These data are presented in input to the MLP; the input layer is structured in such a way that each pixel belonging to each window ( $F_i$ ) is represented by a node (see Figure 3).



**Figure 3.** Neural network configuration. The input layer is structured in such a way that each pixel belonging to each root window ( $G_{NM}^i$ ) resulting from the contextual feature extraction procedure is represented by an input node. The number of output nodes is expressed by:  $K \times K \times C$ , where  $K \times K$  represents the labelled target element dimensions and  $C$  is the number of classes. The number of hidden nodes in the single hidden layer is given by an empirical rule

For a classification problem in which the target element “to be classified” has dimensions  $K \times K$ , and the output should represent the membership of each pixel in the  $K \times K$  frame to a defined class, the dimension of the output layer is expressed by:  $K \times K \times C$ , where  $C$  is the number of classes.

The output of the network, which is usually interpreted to crisp class assignment according to the *winner-take-all* rule, is softened here considering the value of output neurons directly as classification results, and interpreted as degrees of *possibility* or *compatibility* to classes<sup>30</sup>.

The size of the hidden layer should be determined as a compromise between specialization (more hidden nodes) and generalization (fewer hidden nodes). According to Kannelopoulos and Wilkinson we apply an empirical rule that implements this criteria: fixing the number of hidden neurons may be equal to the largest value resulting from doubling the number of inputs and the number of outputs<sup>32</sup>.

The MLP classifier requires a supervised learning procedure. In our context the supervised training set is constituted by *labelled target-elements (LTE)*. Labelling must be performed considering that classification will produce soft results in terms of gradual membership to classes. Indeed, unlike hard conventional classification, a supervised soft classification requires training data with a quantified degree of membership in the concerned classes, to allow the soft classifier to interpret pattern indeterminacy as membership in more than one class.

Various different methods have been put forward in literature to produce adequate soft training<sup>33</sup>.

Since we are addressing high spatial resolution data and gradual memberships are mainly related to boundary conditions, we adopt a simple interpolation method. In particular the compatibility in the inner part of the training polygons, selected by the expert, is labelled 1, and values decrease as they move towards their boundaries by a linear function. The values obtained express the degree of membership to a given class; this procedure is applied for each class considered.

Formally the training set  $T$  is defined as:

$$T = \{ t_i \mid t_i = G_{N_1}^1, \dots, G_{N_M}^M ; LTE_1, \dots, LTE_C \} \text{ where } G_{N_i}^i \text{ is the root}$$

for each enlarging level,  $F_i$ ,

$LTE_k$  is the labelled target element for each class and

$$0 < LTE(i, j) < 1$$

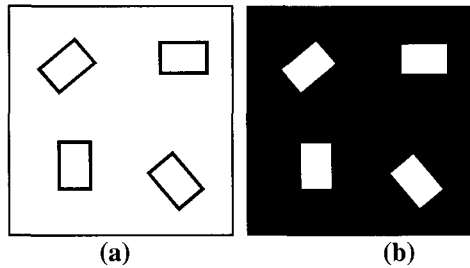
The same sampling procedure is employed for the selection of the testing set to be used in the accuracy assessment phase.

During the generalisation phase the image to be classified is presented in input to the trained network as an arrangement of contiguous target elements,

each associated with the corresponding root images  $G_{N_i}^i$  providing contextual information.

#### 4. Classification of Synthetic Images

A set of experiments were conducted using synthetic images to determine how well the model proposed works in an easily controlled domain.



**Figure 4.** (a)-Synthetic image of 289x289 pixels used in shape-based classification experiment; (b)-Desired classification output

Figure 4-a shows the synthetic image used to determine the capacity of the strategy to extract geometric, shape information, and basing on them to classify objects in the image. Three configurations of the strategy were considered distinguished by an increasing number of enlarging steps in the configuration of the strategy. Figure 5 illustrates the solutions adopted and the corresponding results obtained. In all the three cases the network was trained for 200 epochs and training data have been selected from the image representing the desired output (Figure 4-b) in a proportion of 10% on the total number of pixels.

As seen in this experimental context, the classification benefits from the insertion of more than one level (enlarging step) in the pyramidal feature extraction procedure until the dimensions of the starting window of the level are large enough to avoid ambiguities in proximity to border.

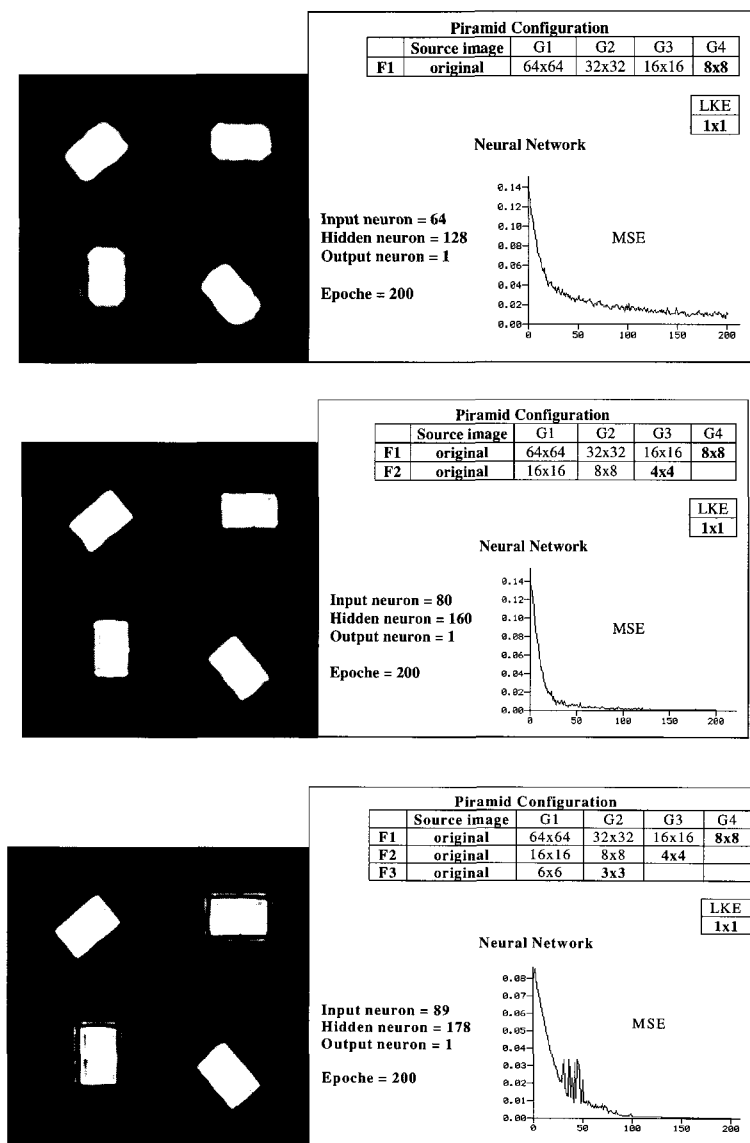
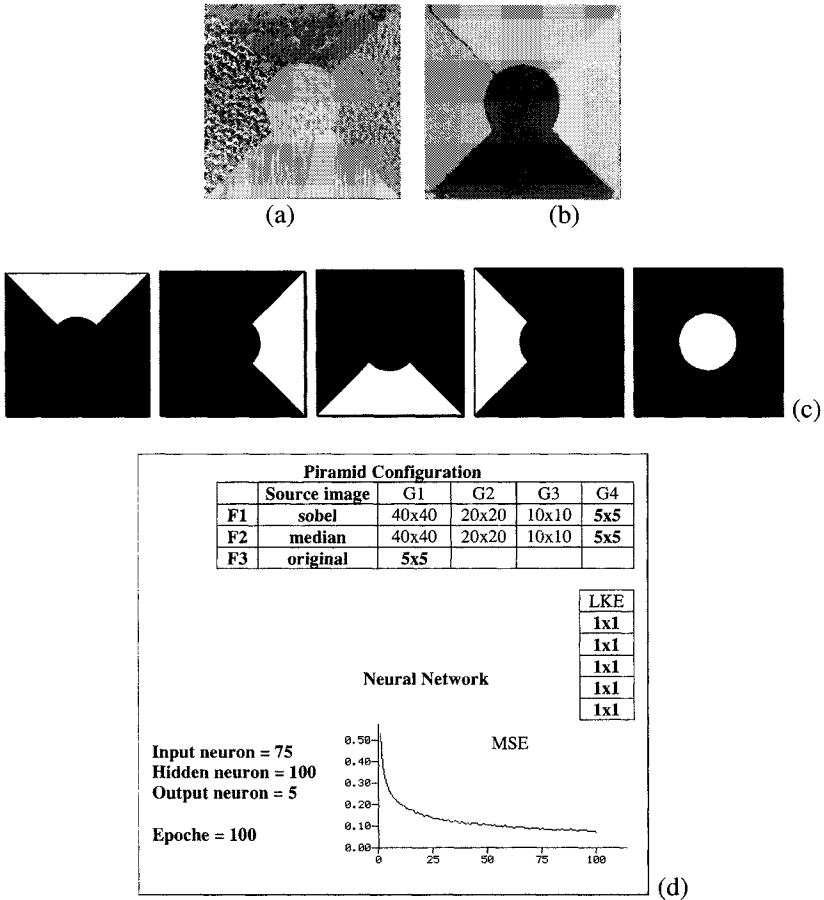


Figure 5. Solutions adopted and results obtained in shape-based classification experiment





**Figure 6.** (a)-Source image composed of 5 natural textures; (b) Classification result (c) Desired classification output; (d) Configuration of the classification strategy

A second experiment has been conducted aimed to develop a texture-based classification. The source image used in the experiment was composed of five natural textures. Figure 6 shows the source and output images together with the configuration of the strategy employed in the experiment. Three enlarging steps have been considered (F1, F2, F3) each with a different image source. F1 and F2 steps have been applied on the image convoluted with Sobel filter to enhance edges before the compacting process; F3 step has been applied on the image resulting from the application of the median filter to the original image to facilitate the blob object detection. The network was trained for 100 epochs and training data have been selected from the image representing the desired output (Figure 6-c) in a proportion of 10% on the total number of pixels.

Results shown in Figure 6-b were encouraging considering the complexity inherent in the texture-based discrimination task.

## 5. Object Recognition for Power Transmission Line Surveillance

The method has been applied to recognition of power transmission lines structures – i.e. towers and lines. The study area was located in Northern Italy, corresponding to the northern part of the second largest basin of the Subalpine lacustrine district: Lake Maggiore. The landscape is typical subalpine: dominated by mixed forest, associated to pastures in the most elevated areas, and to small lawn patches or cultivated areas as moving forward to the lake.

Experimental data were taken from the original  $5005 \times 8993$  IKONOS image set, including: Panchromatic product (1-meter pixel size - acquired at 11bit and processed as remapped at 8 bit), Multispectral (red, green, blue, NIR at 4 meter resolution) and Pan-Sharpned (NIR, green, red at 1 meter resolution) products.

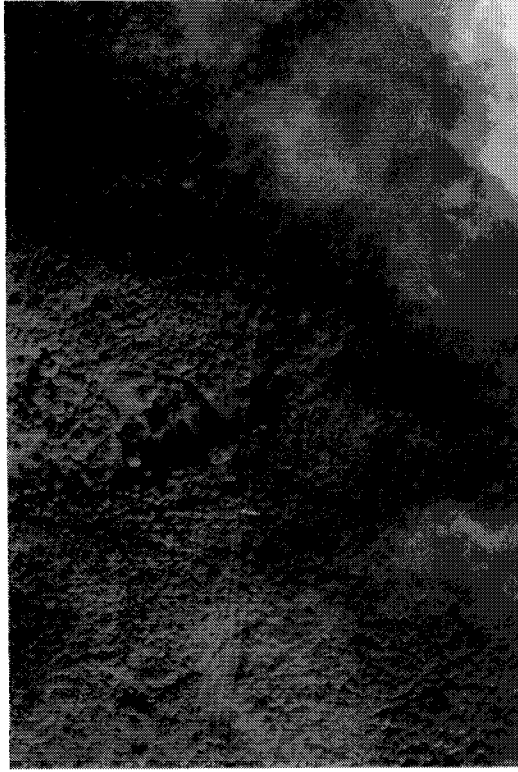
### 5.1 Recognition of towers

The experiments were conducted on a particular typology of transmission towers, those related to 380kV lines (Figure 7).

Figure 8 reports the image sources considered for positioning the concentric windows. Other than Red and Green planes, an additional image plane - obtained by a Sobel filter - was introduced to provide an enhancement of details and to balance the low pass effect of resampling. The Red and Green images were scaled to 1 meter resolution by bicubic resampling

Unfortunately only three towers can be observed in the overall image and the experiment has been conducted taking into account the low availability of target objects. The classifier processes the same objects in learning and allocation phase but with testing areas considerably larger than training areas. Omission errors was interpreted as learning errors and commission errors as generalisation ones.

Figure 9 shows the configuration of the pyramidal model (a) together with desired (b) and obtained results (c).



**Figure 7** Sub-image taken from the original 5005 X 8993 IKONOS multi-spectral product

### *5.2 Recognition of lines*

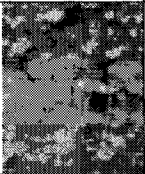
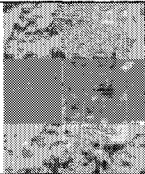
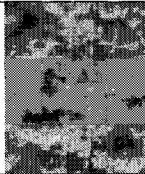



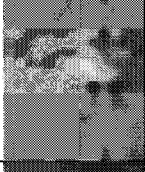
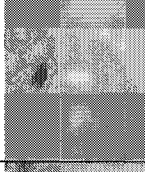
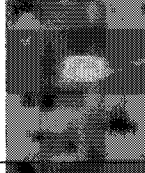
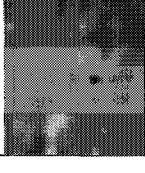
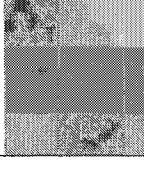
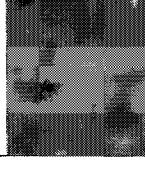
Lines recognition is a support for the recognition of towers. In fact, the superimposition of the of the two classification results - lines and towers - would allow to discard all towers which do not intersect lines. On IKONOS imagery only 380kV lines are detectable. The pyramidal set is configured with two levels:  $3 \times 11$  and  $16 \times 16$  from the pancromatic and red channels respectively (Figure 10).

In Figure 11 the desired and obtained outputs are reported.

To improve classification results the strategy had been applied a second time using as inputs the first classification results and the original red channel. In Figure 12 the strategy configuration as well as results obtained are reported.

Experiments conducted with real imagery, confirmed the feasibility of an automatic detection of power transmission lines structures on IKONOS imagery. The classification strategy, based on contextual feature extraction and neural networks, proved to be an useful tool, either for its flexibility in detecting

different objects (towers, lines) and for the operational level achieved (few parameters for the learning phase configuration, low computational charge).

	Tower 1	Tower 2	Tower 3
<b>Panchromatic</b> 1-meter resolution			
<b>Sobel</b> 1-meter resolution			
<b>Red</b> Scaled to 1-meter resolution			
<b>Green</b> Scaled to 1-meter resolution			

**Figure 8** Image sources considered for extracting contextual information by cognitive pyramid

## 6. Conclusions

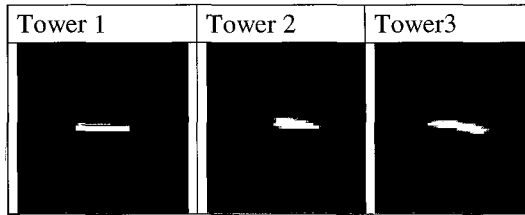
The present work illustrates methodological and application results in the field of high resolution remote sensing object recognition. The contextual classification strategy proposed is considered a cognitive process capable of exploiting and integrating spatial and spectral information for object recognition in remote sensing images. Feature extraction is based on the use of concentric multi-windows and a Gaussian pyramid in an attempt to mimic human perceptual procedures in identifying objects of different shape and structure. Classification is performed by a multi-layer perceptron acting as a soft classifier to deal with unavoidable uncertainty in class discrimination.

As seen in our experimental contexts, recognition benefits from the combined use of multi-resolution feature extraction and neural soft

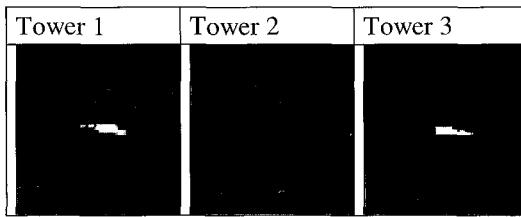
classification. This combination produced an agile and flexible model for high accuracy discrimination of objects that vary in shape, size and structure.

		5x5	Panchromatic
	10	5x5	Sobel
12	6	3x3	Red
12	6	3x3	Green

(a)

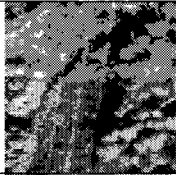
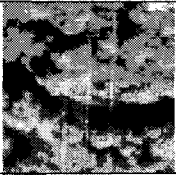




(b)




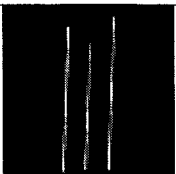
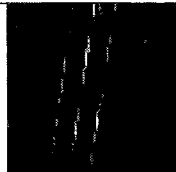
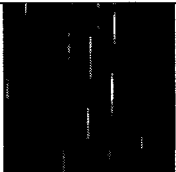
(c)

**Figure 9** Configuration of the model (a) together with desired (b) and obtained results (c)

	fili1	Fili2
<b>Pan</b>		
<b>Red</b>		

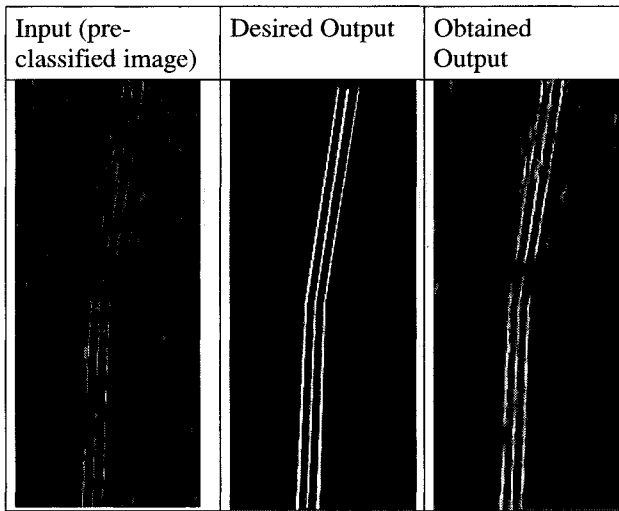
<b>3x11</b>	<b>Pan</b>
<b>16X16</b>	<b>Red</b>

**Figure 10** Image sources and configuration of the model for lines recognition

<b>Lines 1</b>	<b>Lines 2</b>
	
<b>Lines 1</b>	<b>Lines 2</b>
	

**Figure 11** Desired and obtained results in lines recognition

<b>20x20</b>	<b>10x10</b>	<b>5x5</b>	Classified image Red
	<b>10x10</b>	<b>5x5</b>	



**Figure 12** Re-classification results in line recognition experiment

## References

1. B.A. St-Onge and F. Cavayas, *Remote Sens. Environ.*, 61, 82, (1997).
2. J. A. Richards, *Remote Sensing Digital Image Analysis*, Berlin: Springer-Verlag, Berlin, (1993).
3. P. Gong, and P.J. Howarth, *Photogrammetric Engineering and Remote Sensing*, 58, 423, (1992).
4. M.J. Barnsley and S.L. Barr, *Photogrammetric Engineering and Remote Sensing*, 62, (1996).
5. D. Sveinsson J.A. Benediktsson, in *Geospatial Pattern Recognition*, Binaghi. Brivio, Serpico (eds.), Research Signpost, Kerala India, In press.
6. M.E. Hodgson, *Photogrammetric Engineering and Remote Sensing*, 64, 797, (1998).
7. R.A. Schowengerdt, *Remote Sensing-Models and Methods for Image Processing*, Academic Press, London, (1997).
8. P.J. Burt and E.H. Adelson, *IEEE Transactions on Communications*, 31 532, (1983).
9. Mallat S.G, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 11, (1989).

10. Chen C.H. and Du Y., in *Information Processing for Remote Sensing*, C.H. Chen (ed) , World Scientific Publishing, Singapore (1999).
11. Bruce L.M. and Li J., *IEEE Trans. on Geoscience and Remote Sensing*, 39, 1540, (2001).
12. C.M. Gurney and J.R.G. Townshend, *Photogrammetric Engineering and Remote Sensing*, 49, 55, (1983).
13. R.L. Kettig and D.A Langrabe., *IEEE Trans. On Geoscience Electronics*, 14 (1),19, 1976.
14. P. Gong, and P.J. Howarth, *Photogrammetric Engineering and Remote Sensing*, 58, 423, (1992).
15. D.J. Marceau, P.J. Howarth, JM. Dubois and D.J. Gratton, *IEEE Transactions on Geoscience and Remote Sensing*, 28, 513, (1990).
16. C.M. Gurney, *International Journal of Remote Sensing*, 3, 51, (1982).
17. P. Gong and P.J. Howarth, *Remote Sensing of Environment*, 30, 33, (1989).
18. C.C. Kontoes and D. Rokos, *International Journal of Remote Sensing*, 17, 3093, (1996).
19. K.E. Kim, *International Journal of Remote Sensing*, 17, 1083, (1996).
20. A.M. Flygare, *International Journal of Remote Sensing*, 18, 3835, (1997).
21. M.J. Barnsley and S.L. Barr, *Photogrammetric Engineering and Remote Sensing*, 62, 949, (1996).
22. G. Mehltau and R.A. Schowengerdt, *Photogrammetric Engineering and Remote Sensing*, 56, 887, (1990).
23. J.D. Paola and R.A. Schowengerdt, *International Journal of Remote Sensing*, 16, 3033, (1995).
24. E. Binaghi, P. A. Brivio, I. Gallo, M. Pepe, A. Rampini, in *Proceedings of Europto Symposium, SPIE vol. 4170 - Image and Signal Processing for Remote Sensing*, 49, (2001).
25. P. Gong., D.J. Marceau and P.J. Howarth, *Remote Sensing of Environment*, 40, 137, 1992.
26. J.R. Jensen. *Introductory Digital Image Processing: a Remote Sensing Perspective*. Second edition, Prentice-Hall, Englewood Cliffs, New Jersey, (1996).
27. J.S. Greenfield, *Photogrammetric Engineering and Remote Sensing*, 57 1049, (1991).
28. R.A. Schowengerdt. *Remote sensing: models and methods for image processing*. Second edition, Academic Press, San Diego, CA, USA, 522 (1997).
29. P.J. Burt, *Computer Graphics, Image Processing*, 16,20, 1981.
30. E. Binaghi, P.A. Brivio., P. Ghezzi, A. Rampini, E. Zilioli, *Canadian Journal of Remote Sensing*, 25, 171, (1999).
31. A. C. Bernard, I. Kannelopoulos, and G. G. Wilkinson, in E. Binaghi, P.A. Brivio, A. Rampini (eds.), *Soft Computing in Remote Sensing Data Analysis, Series in Remote Sensing, Vol. 1*, Singapore: World Scientific Publishing., 53, (1996).



32. I. Kanellopoulos and G. G. Wilkinson, *International Journal of Remote Sensing*, 18, 711, (1997).
33. G.M Foody, *International Journal of Remote Sensing*, 17, 1317, (1996).

# **Change Detection**

This page is intentionally left blank

## CHAPTER 18

### ADVANCES IN UNSUPERVISED CHANGE DETECTION

Gabriele Moser

*Department of Biophysical and Electronic Engineering  
Via Opera Pia 11a, I-16145 Genoa, Italy  
E-mail: gemini@dibe.unige.it*

Farid Melgani

*Department of Information and Communication Technology  
Via Sommarive 14, I-38050, Trento, Italy  
E-mail: melgani@dit.unitn.it*

Sebastiano B. Serpico

*Department of Biophysical and Electronic Engineering  
Via Opera Pia 11a, I-16145 Genoa, Italy  
E-mail: vulcano@dibe.unige.it*

The analysis of remote-sensing images has proved to be a powerful tool for monitoring the state of the Earth surface in several applications. Consequently, the availability of accurate and reliable algorithms for the detection of changes in remote sensing images is an important issue of growing interest. After a brief survey of previous work in change detection (including unsupervised, supervised and partially supervised approaches), two new unsupervised techniques, recently proposed by the authors, are described. The first is based on a modification to a thresholding algorithm originally proposed in the context of computer vision applications. The second involves the combination of the Fisher transform with the Expectation-Maximization algorithm. Experimental results on both simulated and real data sets and a comparison with another unsupervised approach are presented and discussed.

#### 1. Introduction

The development of effective methodologies and efficient tools for the analysis of multitemporal remote-sensing images represents one of the most important challenges that the remote-sensing community will face in the next years. In this context, a strategic role is played by approaches devoted to detecting changes in images acquired in the same geographical area but at different times. The growing inter-

est in change detection is mainly motivated by two reasons. The first is related to the large number of potential applications of change-detection methods, like environmental monitoring,<sup>1</sup> urban studies,<sup>2</sup> hazard monitoring,<sup>3</sup> forest monitoring,<sup>4</sup> etc. The second is associated with the characteristics of recent and future Earth-observation sensors, which aim to sharply reduce the revisit time from weeks to days or less, thus making it possible to realize several new applications related to change detection.

Change detection in multitemporal remote-sensing images is characterized by several peculiar factors that render ineffective most of the multitemporal image-analysis techniques typically used in other application domains. The main difficulties with change detection in remote-sensing images arise from: the lack of *a-priori* information about the shapes of changed areas; the absence of a reference background; differences in light conditions, atmospheric conditions, sensor calibration, and ground moisture at the acquisition dates; problems of registration of multitemporal images.<sup>5,6,7,8,9</sup> These factors restrict the use of most classical multitemporal image-analysis techniques to few particular remote-sensing applications and make change detection in multitemporal remote-sensing images a challenging problem requiring *ad hoc* methodologies.

Generally, two main kinds of objectives can be pursued in change detection problems: 1) detection and location of the changes that have occurred in a pair of images between the related acquisition dates; 2) identification of typologies of land-cover transitions. The former objective is usually attained in a completely unsupervised way, whereas the latter is generally accomplished in a supervised or partially supervised way.

In the present chapter, after a brief survey of existing change-detection techniques for remotely sensed imagery, we describe two unsupervised change-detection methodologies. The first methodology is based on a thresholding algorithm<sup>35</sup> originally developed in the context of computer vision applications. The second methodology represents a modification to a recently proposed unsupervised<sup>14</sup> technique based on the Expectation-Maximization (EM) algorithm. In particular, we aim to apply EM in a transformed space, computed by the Fisher transform. In addition, we propose to exploit the improved version of EM developed by Jackson *et al.*<sup>42</sup> For experimental investigations, a study area damaged by a forest fire is considered. Two Landsat TM images of the area (acquired before and after the event) are utilized to detect the burnt zones and to assess the proposed techniques and compare them with the above unsupervised approach.<sup>14</sup>

## 2. Previous work

Any change-detection problem can be defined as a classification problem where a "change" class and a "no-change" class have to be distinguished, given two input images. In order to generate change/no-change maps, completely unsupervised approaches are generally used, as they do not require the "expensive" process of

ground truth collection. Therefore, no prior information about the statistics of the aforesaid classes is available to the classification algorithm. Actually, the application of any unsupervised change detection method to two registered (possibly multispectral) images of the same area does not directly detect changes in the land cover; instead, it detects changes in the radiance that reaches a sensor. Therefore, unsupervised change detection algorithms generally imply a fundamental assumption, i.e., variations in radiance due to land-cover changes are greater than those due to sensor noise, different atmospheric conditions or different soil moisture, etc.<sup>10</sup> Under this assumption, a significant variety of unsupervised change detection methods has been proposed, such as image differencing, image regression, image ratioing, change vector analysis (CVA), and methods based on principal component analysis (PCA) or on the analysis of band ratios and vegetation indexes.<sup>10</sup> One of the most widely used approaches is image differencing, according to which the images acquired at two dates are subtracted in order to produce a "difference image" to be analyzed. When input data are multispectral, the CVA technique, which subtracts the two (vectorial) images in order to produce a "spectral change vector" image, can be applied. Changes are identified by computing and analyzing the modulus of this change vector, whereas its direction gives information about the kind of changes. Both the difference image and the modulus image have to be processed in some way, in order to obtain a final change map. The underlying idea is that no-change pixels exhibit small values in the modulus image, whereas change areas exhibit larger values. Therefore, a simple thresholding algorithm can be used in order to distinguish between these two possibilities. However, the thresholding approach involves a further problem, i.e., the choice of an appropriate threshold, which should be optimized according to some criterion for the correct detection of changes.

An alternative to the thresholding method for the solution of an unsupervised change-detection problem is the unsupervised classification approach, which assumes a given mathematical model for the statistics of the "change" and "no-change" classes, performs some kind of estimation of the model parameters, and uses these estimates to apply a classification or detection procedure (e.g., MAP, Neyman-Pearson, etc.<sup>11,12,13</sup>). For example, Bruzzone and Prieto<sup>14</sup> proposed a classification method based on the analysis of the modulus of the multispectral difference image, assuming a Gaussian model for both the "change" and "no-change" classes. Therefore, the parameters to be estimated were the means, the variances and the prior probabilities of the distributions of the modulus for the two classes. These estimations were performed by using the Expectation-Maximization (EM) algorithm, which is an iterative procedure converging, under appropriate assumptions, to maximum likelihood (ML) estimates of the unknown parameters.<sup>15,16</sup> In Aach *et al.*<sup>17</sup> the change-detection problem is viewed as an inverse ill-posed problem that is solved by using prior knowledge of typical properties of change masks, described as samples derived from two-dimensional Markov Random Fields (MRFs). Bruzzone and Prieto<sup>18</sup> developed an MRF-based adaptive semiparametric technique that makes

use of the Reduced Parzen Estimate (RPE) and the Expectation-Maximization (EM) algorithm to estimate (in an unsupervised way) the changes that may occur in a temporal sequence of images. In Kasetkasem *et al.*<sup>19</sup>, the observed temporal images are modeled as MRFs in order to search for an optimal image of changes by means of the maximum *a posteriori* probability (MAP) decision criterion and the Simulated Annealing (SA) energy minimization procedure. Finally, in Melgani *et al.*<sup>20</sup>, a comparative study between several basic thresholding algorithms developed for pattern-recognition or computer-vision problems (like image binarization) and more sophisticated thresholding methods based on the EM algorithm is made in the context of change detection in remote-sensing images.

When an application requires the identification of the kinds of land-cover transitions that have occurred, it becomes necessary to integrate the prior knowledge about the study area (ground truth) into the image-analysis process. This can be done in a supervised or a partially supervised way. In the former way, it is assumed that the ground truth is available for all acquired temporal images. In comparison with a completely unsupervised method, such an approach usually leads to more accurate and more reliable results, especially if one uses advanced multitemporal methods capable to take into account temporal correlation between images. We note that any technique for the classification of multitemporal images that allows changes to occur between acquisition dates fall into this category. One of the early classifier is the "cascade" classifier introduced by Swain,<sup>21</sup> which extends the Bayes criterion for the minimum error to the case of multitemporal data. The stochastic model-based approach proposed by Kalayeh *et al.*<sup>22</sup> uses a stochastic system representing land-cover types through a non-stationary Gaussian process as input and the temporal spectral behavior as output. An iterative technique, derived from the Bayes rule for the minimum error, was proposed by Bruzzone *et al.*<sup>23</sup> to estimate the matrix of the probabilities of transitions (TPM). In Bruzzone *et al.*<sup>24</sup>, the TPM is replaced with the Joint Probability Matrix (JPM), computed by applying a specific formulation of the Expectation-Maximization algorithm. Two techniques, based differently on the notion of the minimum expected cost and dealing with the multitemporal aspect at the decision level, were presented by Jeon *et al.*<sup>25</sup>. In Tian *et al.*<sup>26</sup>, a temporal updating approach to cloud classification is developed. In order to track temporal changes in a sequence of images, a Probabilistic Neural Network (PNN) classifier is updated using the maximum likelihood criterion. The use of spatial information to further improve multitemporal classification accuracy was also adopted in the literature. In this context, an interesting (but computationally expensive) autocorrelation model was proposed by Khazenie *et al.*<sup>27</sup> Jeon *et al.*<sup>28</sup> developed a cascade spatio-temporal contextual classifier that involves Markov Random Fields to model spatial correlation. A general model using Markov Random Fields (MRFs) to deal with a sequence of temporal images was proposed by Solberg *et al.*<sup>29</sup> This model is of great interest for its ability to easily integrate different kinds of information (e.g., derived from a GIS), in addition to the temporal one,

into a scene model by using the concept of energy functions. In Melgani *et al.*<sup>30</sup>, a simple statistical model was proposed to merge spectral information with spatial and temporal contextual information. In Yang *et al.*<sup>43</sup> a change-detection method is developed, which integrates clustering information with several typologies of reference data through interactive photointerpretation, under the assumption that no new classes appear between the considered acquisition dates. In Prieto *et al.*<sup>32</sup>, the requirements for ground-truth availability at the two dates were partially relaxed, in the sense that the ground truth was necessary only for the land cover transitions of interest to the end-user. In this context, a partially supervised technique based on Markov Random Fields was proposed that exploits training samples belonging exclusively to the land covers involved in the specific kind of changes to be mapped.

Especially when a repetitive detection of changes is needed to monitor a given geographical area, on the basis of periodical acquisitions of new images, the availability of ground-truth information at all acquisition dates appears to be too onerous. In contrast to the supervised approach, detection of land-cover transitions by the partially supervised approach is based on the more realistic assumption that the ground truth is available for at least one of the multitemporal images. Despite this is a crucial issue in developing effective methods to be used in real applications, in the scientific literature scant attention has been devoted to such an issue. In Bruzzone *et al.*<sup>31</sup> a partially supervised methodology was proposed able to update the parameters of an already trained parametric maximum-likelihood (ML) classifier whenever a new image lacking the corresponding ground truth has to be analyzed. Under the assumption that no new class is contained in the new image, the updating is performed by means of the EM algorithm, which allows one to tune the parameters of the trained ML classifier on the basis of the distribution in the new image. In this way, it is possible to classify multitemporal data of a given area and hence to obtain land-cover transition maps. Moser *et al.*<sup>33</sup> proposed a partially supervised change-detection scheme based on the exploitation of the ground truth availability for at least one temporal image. Considering the case of two images acquired at different dates, they applied first a clustering algorithm to both images, and then a thresholding-based unsupervised change-detection algorithm<sup>35</sup> to each separate cluster of the second image. As a result, their approach not only made it possible to identify the presence of changes but also to distinguish the different typologies of changes.

### 3. A thresholding approach to unsupervised change detection

The thresholding approach<sup>34</sup> to the classification problem aims at splitting the pixels of a scalar image  $\mathcal{Z} = \{z_{mn} : m = 0, 1, \dots, M - 1, n = 0, 1, \dots, N - 1\}$  into two classes, namely,  $\omega_1$  and  $\omega_2$ , by fixing a threshold value  $T \in \mathbb{R}$  and generating a



classification map  $\mathcal{L} = \{\ell_{mn} : m = 0, 1, \dots, M - 1, n = 0, 1, \dots, N - 1\}$  such that:

$$\ell_{mn} = \begin{cases} \omega_1 & \text{if } z_{mn} \leq T \\ \omega_2 & \text{if } z_{mn} > T \end{cases} \quad (1)$$

This approach has been employed for computer-vision classification problems,<sup>35,36,37</sup> for instance, for separation of the "foreground" from the "background". In the present chapter, we consider thresholding as a solution to change-detection problems for remotely sensed multitemporal imagery in order to separate the "change" class from the "no-change" class in the difference image. The problem of how to transform the difference image (which is vectorial) into a scalar image will be addressed later on.

In particular, we assume to operate in a purely unsupervised context, characterized by a complete lack of prior knowledge about the classes under consideration. In a supervised context, the selection of a suitable threshold value, aimed at minimizing the error probability in the binary classification map, would be simply performed by applying the Bayes decision theory,<sup>34</sup> after estimating the class statistics on the available training set. In the present work, an unsupervised threshold-selection method is instead proposed.

### 3.1. Kittler and Illingworth's thresholding method

For the unsupervised change-detection problem under investigation, we adopt an unsupervised threshold selection approach, originally proposed for artificial vision purposes. The method assumes the input scalar image  $\mathcal{Z}$  to present  $L$  possible grey levels (namely,  $\{0, 1, \dots, L - 1\}$ ) and the image histogram  $h(z)$  ( $z = 0, 1, \dots, L - 1$ ) to be the only available information about the image itself (i.e., the method uses only grey-level information). The class-conditional probability density functions (pdfs)  $p_i(z) \triangleq p(z|\omega_i)$  ( $i = 1, 2$ ) of the grey level  $z$  are not known in advance, nor the prior probabilities  $P_i \triangleq P(\omega_i)$  ( $i = 1, 2$ ) of the classes. As a consequence, in place of the pdf:

$$p(z) = P_1 p_1(z) + P_2 p_2(z) \quad (2)$$

obtained by the total probability theorem,<sup>13,39</sup> the histogram  $h(z)$  itself is computed and utilized.

The selection of an appropriate threshold  $T$  is based on the optimization of a given predefined criterion function  $J(T)$  that averages a cost function  $c(z, T)$  with respect to the feature  $z$ . This means that the threshold is assumed to be fixed at some value  $T \in \{0, 1, \dots, L - 1\}$ , and that a cost function  $c(z, T)$  is introduced that measures the cost of classifying pixels by comparing their grey levels  $z$  with the threshold  $T$ . The criterion function is then computed by averaging over the feature

histogram  $h(z)$ :

$$J(T) = \sum_{z=0}^{L-1} h(z)c(z, T) \quad (3)$$

and the optimal threshold that minimizes the criterion function is defined as follows:

$$T^* \triangleq \arg \min_{T=0, \dots, L-1} J(T) \quad (4)$$

We have considered several algorithms that can be formalized within this framework, depending on the specific model adopted for the cost function.<sup>34,20</sup> Otsu's method<sup>37</sup> defines a cost function performing a discriminant analysis of the image histogram; Kittler and Illingworth's algorithm<sup>35</sup> applies the Bayes decision theory to derive a minimum-error criterion; and Huang and Wang's procedure<sup>36</sup> optimizes a fuzzy-based criterion function. In this chapter, we focus our attention on Kittler and Illingworth's algorithm<sup>35</sup>, as, in a previous comparative analysis of the considered methods,<sup>20,38</sup> it obtained very good accuracy results, thus proving to be an effective unsupervised classification tool and to perform much better than the other two algorithms.

Kittler and Illingworth define a cost function according to the MAP (maximum *a posteriori*) classification rule, under the Gaussian assumption for the class-conditional probabilities (i.e.,  $p_i(z) = \mathcal{N}(m_i, \sigma_i^2)$  for  $i = 1, 2$ ). Under this hypothesis, the only parameters to be estimated are the class prior probabilities  $P_1$  and  $P_2$ , the class-conditional means  $m_1$  and  $m_2$ , and the class-conditional variances  $\sigma_1^2$  and  $\sigma_2^2$ . Given the threshold value  $T$ , the following estimates are introduced:

$$\hat{P}_1(T) \triangleq \sum_{z=0}^T h(z), \quad \hat{P}_2(T) \triangleq \sum_{z=T+1}^{L-1} h(z) \quad (5)$$

$$\hat{m}_1(T) \triangleq \frac{1}{\hat{P}_1(T)} \sum_{z=0}^T zh(z), \quad \hat{m}_2(T) \triangleq \frac{1}{\hat{P}_2(T)} \sum_{z=T+1}^{L-1} zh(z) \quad (6)$$

$$\hat{\sigma}_1^2(T) \triangleq \frac{1}{\hat{P}_1(T)} \sum_{z=0}^T [z - \hat{m}_1(T)]^2 h(z), \quad \hat{\sigma}_2^2(T) \triangleq \frac{1}{\hat{P}_2(T)} \sum_{z=T+1}^{L-1} [z - \hat{m}_2(T)]^2 h(z) \quad (7)$$

Hence, the estimates  $\hat{p}_i(z|T)$  ( $i = 1, 2$ ) of the Gaussian class-conditional pdfs  $p_i(z)$  are implicitly defined, and the estimates  $\hat{P}^{(i)}(z, T)$  of the posterior probabilities  $P^{(i)}(z) = P(\omega_i|z)$  of the grey level  $z$  belonging to the class  $\omega_i$  ( $i = 1, 2$ ) are computed by applying the Bayes rule.<sup>13,39</sup> In order to penalize incorrect pixel classification, Kittler and Illingworth define their cost function as follows:

$$c(z, T) \triangleq \begin{cases} -2 \ln \hat{P}^{(1)}(z, T) & \text{if } z \leq T \\ -2 \ln \hat{P}^{(2)}(z, T) & \text{if } z > T \end{cases} \quad (8)$$

Except for an additive constant, the result is

$$c(z, T) = \begin{cases} \frac{[z - \hat{m}_1(T)]^2}{\hat{\sigma}_1^2(T)} - 2 \ln \frac{\hat{P}_1(T)}{\hat{\sigma}_1(T)} & \text{if } z \leq T \\ \frac{[z - \hat{m}_2(T)]^2}{\hat{\sigma}_2^2(T)} - 2 \ln \frac{\hat{P}_2(T)}{\hat{\sigma}_2(T)} & \text{if } z > T \end{cases} \quad (9)$$

which is a pair of parabolic arcs (centered in the class-conditional means), whose shapes depend on the threshold  $T$  (Fig. 1):

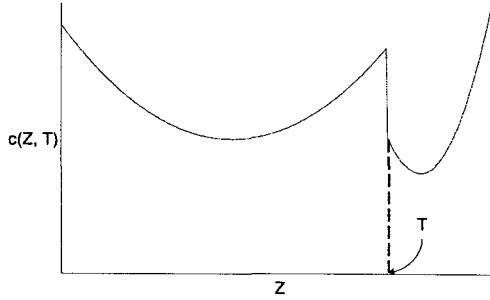


Fig. 1. Plot of the cost function  $c(z, T)$ .

Except for additive and multiplicative constants, the resulting criterion function is proved to be:

$$J(T) = \hat{P}_1(T) \ln \hat{\sigma}_1(T) + \hat{P}_2(T) \ln \hat{\sigma}_2(T) - \hat{P}_1(T) \ln \hat{P}_1(T) - \hat{P}_2(T) \ln \hat{P}_2(T) \quad (10)$$

In the following, we shall refer to Kittler and Illingworth's algorithm simply as K&I .

### 3.2. Unsupervised change detection by the K&I algorithm

As thresholding is *per se* a scalar operator (i.e., the thresholding operator accepts only scalar images as input), it cannot be directly applied to the difference image, which is typically multispectral. Therefore, we propose to use the K&I method after a preliminary  $d$ -to-1 feature transform from the  $d$ -dimensional space of the difference-image bands into a one-dimensional transformed space (Fig. 2). Specifically, for this processing stage we considered two unsupervised transforms, namely, the modulus (i.e., Euclidean norm) operator and the first component of the Principal Component Analysis (PCA).<sup>12,13</sup>

From the linearity of the PCA operator it results that, if the difference image is assumed to be a Gaussian mixture of two components corresponding to the "change" and "no-change" areas, the PCA first component image is Gaussian, too. On the contrary, due to the non-linearity of the Euclidean norm, the modulus of a Gaussian mixture is not a Gaussian mixture any more. This is a drawback of the modulus used

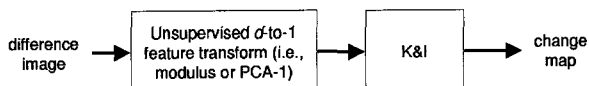


Fig. 2. Block diagram of the unsupervised change-detection method based on K&I .

as an unsupervised transform for the proposed algorithm, as K&I assumes the class-conditional normality of the input image. On the other hand, the PCA transform is data-dependent (as it is computed by diagonalizing the sample-covariance matrix of the image), whereas the modulus is not. Hence, the PCA first component image may be affected by more deviations than the assumed two-component Gaussian mixture model. In the following, this transformed feature will be simply denoted by PCA-1. K&I applied to the modulus or to the PCA-1 image is the first technique that will be experimentally investigated.

#### 4. Classification-based approach to unsupervised change detection

An alternative to the thresholding method for the solution of an unsupervised change-detection problem is an unsupervised parametric classification approach, which postulates a given mathematical model for the statistics of the "change" and "no-change" classes and estimates the parameters of the model.

In the present unsupervised context, also the parameter-estimation stage must be unsupervised. A feasible method is the Expectation-Maximization (EM) algorithm, which can be applied to any estimation problem affected by some kind of incompleteness of the available data,<sup>15,16</sup> that is, when it is not possible to have access directly to the complete data but only to a set of incomplete data. Other applications of EM can be found in the literature,<sup>16</sup> such as active noise cancellation, hidden Markov models, spread-spectrum multiuser communication, etc. This large variety of applications stems from the considerable generality of the EM algorithm for the solution of incompleteness problems, and witnesses its effectiveness as an estimation tool.

The approach based on EM followed by classification was proposed by Bruzzone *et al.*<sup>14</sup>. In this paper, we develop two modifications to that approach, one involving a different version of the EM algorithm itself<sup>42</sup>, the other related to the space where the EM-based estimation procedure is to be applied.

##### 4.1. EM-based unsupervised classification

The Expectation-Maximization (EM) algorithm<sup>15,16</sup> provides a general solution to the problem of the estimation of the parameters of a pdf in the presence of incomplete data. It is implemented in a pair of steps, namely, the E-step, which computes a pseudo-likelihood function, and the M-step, which maximizes the pseudo-likelihood itself. These steps are iterated up to convergence,<sup>38</sup> and it has been

demonstrated<sup>15,41</sup> that, under quite general assumptions, the algorithm converges to a maximum-likelihood (ML) estimate of the parameter vector. Even though it may suffer from local minima and saddle points, the EM algorithm has proved to be effective, in particular, for the kind of problems considered here, i.e., estimation of mixture-component parameters.<sup>15,40,42,38</sup>

Focusing on the specific problem of unsupervised classification, given an  $M \times N$  multispectral image  $\mathcal{Z} = \{z_{mn} \in \mathbb{R}^d : m = 0, 1, \dots, M - 1, n = 0, 1, \dots, N - 1\}$  consisting of  $d$  bands, for the pixel-intensity random vector we assume a  $K$ -component Gaussian mixture model, i.e.,

$$p(z) = \sum_{i=1}^K P_i p_i(z) \quad (11)$$

where  $p_i(z) \triangleq p(z|\omega_i) = \mathcal{N}(m_i, \Sigma_i)$  is the pdf of the pixel-intensity vector  $z$  conditioned by the class  $\omega_i$ , and  $P_i \triangleq P(\omega_i)$  is the prior probability of  $\omega_i \in \Omega = \{\omega_1, \dots, \omega_K\}$ . The parameters characterizing this statistical model are the prior probabilities  $\{P_i : i = 1, 2, \dots, K\}$ , the components of the  $d$ -dimensional mean vectors  $\{m_i : i = 1, 2, \dots, K\}$ , and the elements of the  $d \times d$  covariance matrices  $\{\Sigma_i : i = 1, 2, \dots, K\}$ . We collect these parameters in the vector  $\theta$ , and denote explicitly  $p_i(z)$  and  $p(z)$  by  $p_i(z|\theta)$  and  $p(z|\theta)$ , respectively.

It is possible to show that a standard application of the EM algorithm leads to a set of iterative equations such that, at each iteration step  $t$ , all pixel values are used to update the parameters of all classes.<sup>38,42</sup> Jackson and Landgrebe<sup>42</sup> suggested adding an intermediate classification step between the classic E-step and the M-step: at each iteration, the current values of the pdf parameters are used to classify the pixels, and the parameters of a given class are updated considering only the pixels assigned to that class. The use of such "semi-labeled" samples aims at reducing the overlapping between classes, also integrating parameter estimation and classification. Specifically, Jackson *et al.*<sup>42</sup> adopted an ML classifier and made use of a training set. In the present work, we adapt their version of EM iterations to our completely unsupervised context and to the use of a MAP classifier.<sup>38</sup> Thus, at the  $t$ -th iteration, the following operations are performed:

- **Expectation step:** for all  $z \in \mathcal{Z}$  and  $i = 1, \dots, K$ , the  $t$ -th step EM estimate of the posterior probability  $P(\omega_i|z)$  ( $z \in \mathcal{Z}, i = 1, \dots, K$ ) is computed:

$$P(\omega_i|z, \theta^{(t)}) = \frac{P_i^{(t)} p_i(z|\theta^{(t)})}{p(z|\theta^{(t)})} \quad (12)$$

- **Classification step:** according to a MAP classifier, the decision region for the class  $\omega_i$ , i.e., the subset of samples that MAP assigns to the class  $\omega_i$ , is:

$$\mathcal{I}_i^{(t)} = \{z \in \mathcal{Z} : P(\omega_i|z, \theta^{(t)}) \geq P(\omega_j|z, \theta^{(t)}) \quad \forall j \neq i\} \quad (13)$$

- **Maximization step**<sup>a</sup>: for all  $i = 1, 2, \dots, K$ :

$$P_i^{(t+1)} = \frac{\sum_{z \in \mathcal{I}_i^{(t)}} P(\omega_i | z, \theta^{(t)})}{\sum_{j=1}^K \sum_{z \in \mathcal{I}_j^{(t)}} P(\omega_j | z, \theta^{(t)})} \quad (14)$$

$$m_i^{(t+1)} = \frac{\sum_{z \in \mathcal{I}_i^{(t)}} P(\omega_i | z, \theta^{(t)}) z}{\sum_{z \in \mathcal{I}_i^{(t)}} P(\omega_i | z, \theta^{(t)})} \quad (15)$$

$$\Sigma_i^{(t+1)} = \frac{\sum_{z \in \mathcal{I}_i^{(t)}} P(\omega_i | z, \theta^{(t)}) (z - m_i^{(t+1)}) (z - m_i^{(t+1)})^T}{\sum_{z \in \mathcal{I}_i^{(t)}} P(\omega_i | z, \theta^{(t)})} \quad (16)$$

#### 4.2. Fisher/EM unsupervised change-detection method

For change-detection purposes, we assume the multispectral difference image to be a mixture of two Gaussian populations, corresponding to "change" and "no-change" pixels, respectively. The EM algorithm could be applied to estimate the parameters of the two Gaussian distributions directly in the  $d$ -dimensional space of the difference image. However, especially in a completely unsupervised context (i.e., with no training samples), the results may turn out to be more sensitive to deviations from the assumed distributional model. As a consequence, we propose to apply EM in a one-dimensional space. We do not apply EM to the modulus image because the modulus is nonlinear and strongly affects the data normality. On the contrary, linear transformations improve the data Gaussianity,<sup>44</sup> making a parametric Gaussian-mixture model more realistic and more accurate, thus increasing the effectiveness of Eqs. 12-16 in the estimation stage. In particular, we have selected the Fisher transform, as it performs a discriminant analysis, mapping the  $d$ -dimensional feature vector  $z$  of the bands of the difference image into a scalar transformed feature  $v$ :

$$v = w^T z \quad (17)$$

where the weight vector  $w$  is chosen such as to maximize the class separation and to minimize the class dispersion<sup>12,13</sup>. In Melgani *et al.*<sup>20</sup> we have also considered the application of the EM algorithm after transforming the difference image by another linear operator, that is, simultaneous diagonalization.<sup>13</sup> Results on real and simulated data<sup>38</sup> have proved that the Fisher transform is more effective.

The Fisher transform is inherently supervised, i.e., it requires an input training set to estimate class-conditional means and scatter matrices.<sup>12</sup> Therefore, as a first step, we propose to apply Kittler and Illingworth's algorithm to some  $d$ -to-1 dimensional unsupervised transform (e.g., modulus or PCA-1) in order to obtain a first classification map. This map is utilized to generate a training set to compute the

<sup>a</sup>All vectors are assumed to be column vectors and the apex "T" denotes the matrix transpose operator.

Fisher transform image  $\mathcal{V} = \{w^T z_{mn} : m = 0, 1, \dots, M-1, n = 0, 1, \dots, N-1\}$  and to provide the first estimates of the parameters of the Gaussian mixture components in the Fisher domain to initialize the EM algorithm.

In particular, denoting by  $\mathcal{U} = \{u_{mn} : m = 0, \dots, M-1, n = 0, \dots, N-1\}$  the unsupervised transformed image and by  $u$  the corresponding scalar feature, the training set for the "no-change" class  $\omega_1$  is defined fixing a threshold  $T_1$  and selecting the pixels lying in  $\mathcal{U}$  to the left of  $T_1$ . For the "change" class  $\omega_2$ , a second threshold  $T_2$  is introduced and the pixels lying in  $\mathcal{U}$  to the right of  $T_2$  are employed as training samples. Hence, the training sets for the two classes are:

$$\mathcal{D}_1(T_1) \triangleq \{(m, n) : u_{mn} < T_1\}, \quad \mathcal{D}_2(T_2) \triangleq \{(m, n) : u_{mn} > T_2\} \quad (18)$$

Denoting by  $\hat{\mu}_1$  and  $\hat{\mu}_2$  the K&I-based estimates of the class-conditional means  $\mu_1 = E\{u|\omega_1\}$  and  $\mu_2 = E\{u|\omega_2\}$ ,  $T_1$  is computed such that  $\mathcal{D}_1(T_1)$  contains a percentage  $s$  of the number of samples that lie to the left of  $\hat{\mu}_1$  and  $T_2$  is such that  $\mathcal{D}_2(T_2)$  contains the same percentage  $s$  of the samples that lie to the right of  $\hat{\mu}_2$ <sup>b</sup>:

$$|\mathcal{D}_1(T_1)| = s |\mathcal{D}_1(\hat{\mu}_1)|, \quad |\mathcal{D}_2(T_2)| = s |\mathcal{D}_2(\hat{\mu}_2)| \quad (19)$$

As a result, biased estimates are obtained, but the most critical area, where the two distributions overlap, is discarded. In the following, we shall synthetically refer to this methodology as Fisher/EM. A complete block diagram of this change-detection algorithm is shown in Fig. 3.

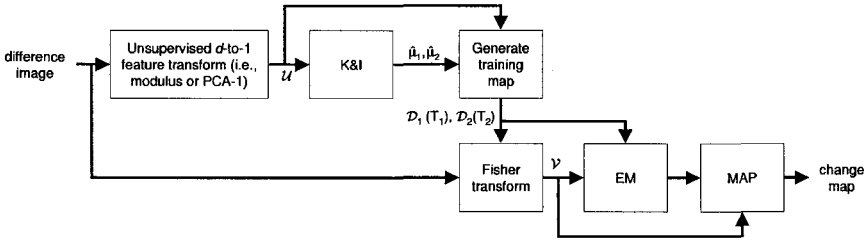


Fig. 3. Block diagram of the Fisher/EM change-detection method.

## 5. Experimental results

### 5.1. Experiments with synthetic data sets

For a preliminary assessment of the potentialities and limitations of the proposed unsupervised techniques, several experiments were performed with two-dimensional synthetic data sets, generated in order to focus on specific issues.<sup>38</sup> We briefly describe here the main experiments and draw the related conclusions.

<sup>b</sup>Given a finite set  $A$ , we denote by  $|A|$  the number of elements (i.e., the cardinality) of  $A$ .

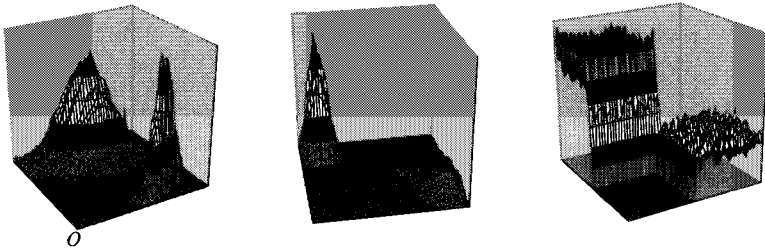


Fig. 4. Two-dimensional histograms for synthetic data sets A (left), B (middle) and C (right).

The first synthetic data set, called "data set A", was generated to simulate a situation in which a fundamental assumption about CVA is not satisfied, i.e, it is not true that low intensity values in the modulus image correspond to "no-change" areas, whereas high values are related to the "change" class. Data set A consists of two well-separated Gaussian classes (Fig. 4). Low error rates were achieved by both proposed methods when using PCA-1 as an unsupervised transform with error probabilities  $P_e = 8.7 \cdot 10^{-4}$  for K&I and  $P_e = 8.1 \cdot 10^{-4}$  for Fisher/EM. On the contrary, very poor results were obtained by employing the modulus image. This failure of the modulus operator is due to the fact that, in the case of the modulus operator, the above fundamental assumption about CVA is critical. In particular, the two-dimensional histogram of data set A (Fig. 4) shows that the two Gaussian modes lie in the feature space at about the same distance from the origin (which is placed at point  $O$  in Fig. 4). Therefore, the required separation in the modulus space is not available for this data set, i.e., the resulting modulus histogram is essentially unimodal. On the contrary, the two classes are well-separated in the PCA-1 domain, as, for this data set, the histogram in the direction of the maximum dispersion (which always corresponds to the first principal component) is bimodal with quite a small overlap.

Data set B (Fig. 4) consists of well-separated normal classes, with independent features. We employed this data set in an experimental study of the stability of the proposed algorithms in the presence of noise. Specifically, classifying data set B is expected to be an easy task for all methods, as both algorithms exhibit no classification errors. Then, starting from this easy situation, we added to both features a Gaussian noise of growing variance, testing the performances of the described methods as a function of the amount of noise. The results show that the error probabilities of K&I and Fisher/EM regularly increase with the variance of the additive noise, without oscillations or non-monotonic behaviors.<sup>38</sup> Hence, in this experiment, K&I and Fisher/EM exhibited regular and stable behaviors in the presence of undesired additive Gaussian noise.

Finally, data set C (Fig. 4) shows a non-standard situation, where the class-conditional pdfs are not Gaussian but uniformly distributed. This data set was generated in order to evaluate the sensitivity to deviations from model statistics,



since both proposed algorithms assume a two-component Gaussian mixture model as a basic hypothesis. Hence, the classification performances are expected to be affected by the inaccuracy of this model. This is confirmed by the results: both proposed methods generated several erroneously classified pixels, despite the good separation between the classes in the feature space. In particular, K&I obtained error probabilities equal to 0.01037 with PCA-1 and equal to 0.005289 with the modulus, whereas Fisher/EM achieved  $P_e = 0.00544$  with PCA-1 and  $P_e = 0.00538$  with the modulus. Therefore, at least for this data set, the most sensitive technique seems to be K&I when applied to the PCA-1 image.

## 5.2. Results on real data

In the experiments with real data, we used a multitemporal data set consisting of a couple of multispectral Landsat 5 Thematic Mapper (TM) images of the western part of the Island of Elba, acquired in August and September 1994. The images show a  $414 \times 326$  portion of a scene representing part of the island and the surrounding sea (the September image was registered to the August one). During the period between the acquisition dates, a fire destroyed a vegetation zone, visible in the considered area. Bands 4 and 7 were used in the experiments, as they are the most effective to detect forest areas damaged by fire.<sup>14</sup> Band 4 and the ground truth image are shown in Fig. 5. Before using the image differencing approach, a simple running mean filter (with a  $3 \times 3$  window) was applied to each of the two considered bands to reduce the noise.

It is important to point out that the difference image did not represent a two-class scene (i.e., "burnt vegetation" and "non-burnt vegetation") but, more realistically, a three-class scene, where "burnt vegetation," "sea," and "non-burnt vegetation" were contained. The presence of three classes (corresponding to three modes in the difference image) can strongly affect the classification accuracies of the previously described algorithms, which refer to bimodal contexts. Therefore, we repeated the experiments also on a reduced portion of the same data set, from which it was obtained by cutting only a  $228 \times 228$  zone of the island in order to keep the "burnt vegetation" and "non-burnt vegetation" classes and to leave out the "sea" class (the cutting window is shown in Fig. 5).

Given the bispectral images (bands 4 and 7) acquired in August and September, the vectorial difference image was computed as well as the modulus and the PCA-1 images (Fig. 6). K&I and Fisher/EM were tested employing both the modulus and the PCA-1 images as input, considering both the  $414 \times 326$  and  $228 \times 228$  versions. Fisher/EM was applied using different values of the related thresholds (i.e., for different choices of the percentage  $s$ ). The resulting accuracies were compared with the ones provided by Bruzzone and Prieto's algorithm.<sup>14</sup> In all the experiments, the probabilities of false alarm, detection, and error (denoted by  $P_F$ ,  $P_D$  and  $P_e$ , respectively) were computed by comparing the output classification maps with the ground truth. The values obtained for both versions of the data set are given in

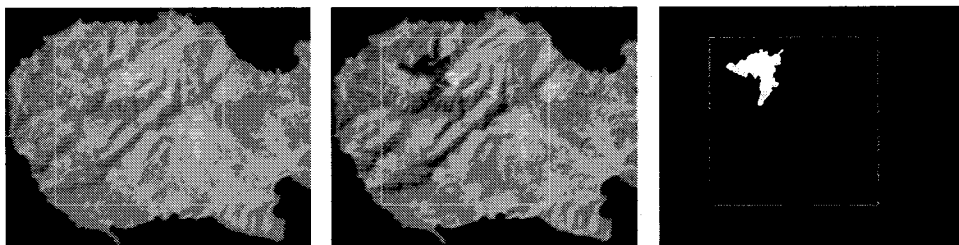


Fig. 5. Multitemporal data set: band 4 of a Landsat-5 TM image, acquired in August 1994 (left); band 4 of a Landsat-5 TM image, acquired in September 1994 (middle); ground truth for the study case with the burnt area represented in white (right). The white square indicates the reduced data set ( $228 \times 228$  pixels in size).

Tables 1-3.

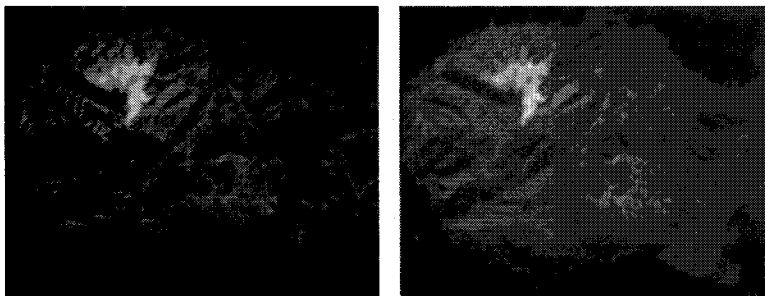


Fig. 6. Modulus (left) and PCA-1 (right) of the difference image (after histogram stretching).

We used Bruzzone and Prieto's algorithm (which was originally developed according to the standard version of the EM iteration scheme) both for a comparison with the proposed methods and for a comparison between the performances of the standard and modified versions of EM. Therefore, for this algorithm only, in our experiments we adopted both versions of EM.

Table 1. Performances of the K&I algorithm applied to the PCA-1 and modulus images.

	228 × 228 data set		414 × 326 data set	
	PCA-1	modulus	PCA-1	modulus
$P_F$	0.00167	0.00135	0.00042	0.00042
$P_D$	0.95568	0.92088	0.86993	0.91135
$P_e$	0.00366	0.00496	0.00273	0.00200

Table 2. Performances of the Fisher/EM algorithm, with the PCA-1 image or the modulus image as unsupervised transforms:  $P_F$ ,  $P_D$  and  $P_e$  are the error probabilities obtained for  $s = 0.6$ ;  $\text{std}(P_e)$  is the standard deviation of  $P_e$  for  $s$  in the range  $[0.2, 1]$ .

	228 × 228 data set		414 × 326 data set	
	PCA-1	modulus	PCA-1	modulus
$P_F$	0.00190	0.00178	0.00099	0.00125
$P_D$	0.96189	0.96230	0.96935	0.97225
$P_e$	0.00358	0.00344	0.00152	0.00172
$\text{std}(P_e)$	0.00018	0.00024	0.00021	0.00034

Table 3. Performances of Bruzzone and Prieto's algorithm applied to the modulus image.

	228 × 228 data set		414 × 326 data set	
	standard EM	modified EM	standard EM	modified EM
$P_F$	0.00464	0.00135	0.00179	0.00042
$P_D$	0.97142	0.92046	0.97349	0.91052
$P_e$	0.00575	0.00498	0.00223	0.00202

All the compared methodologies proved to be able to yield very accurate results. In particular, Fisher/EM provided the best overall classification performances (Table 2) and proved to be quite insensitive to the choice of the initial thresholds  $T_1$  and  $T_2$  for large variations in the parameter  $s$ , that is, from 0.2 to 1. Specifically, the highest accuracies were obtained for  $s \in [0.3, 0.6]$ , whereas a slight increase in the error probability occurred at the extremes of the range. Table 2 provides the performance parameters  $P_F$ ,  $P_D$ , and  $P_e$  for  $s = 0.6$  and the standard deviations of  $P_e$  for values of  $s$  in the range 0.2 to 1. As an example, one of the Fisher-transformed images is shown in Fig. 7.



Fig. 7. Fisher-transformed image (initialization with PCA-1 and  $s = 1$ ).

Also K&I obtained very good results on both versions of the data set, although at slightly higher error rates. In particular, in the 2-mode experiment, K&I attained better results when applied to the PCA-1 image than to the modulus image, whereas in the 3-mode experiments it achieved higher performances when applied to the modulus image. The reason for this difference can be deduced from Fig. 8, which displays the positions of the PCA transform axes in the feature space. It is evident that, in the 2-mode context, projecting in the PCA first component direction  $e_1$  ensures a good class separation. In the 3-mode data set, the direction  $e_1$  is influenced by the presence of the third mode and is almost coincident with one of the original axes, thus discarding a large part of the information conveyed by the other feature, and resulting in a reduced separation between the "change" and "no-change" classes. Hence, in the 3-mode experiment, better results are obtained with the modulus operator, which is data-independent. As an example, Fig. 9 shows the change map and the criterion function provided by K&I for the  $414 \times 326$  modulus image.

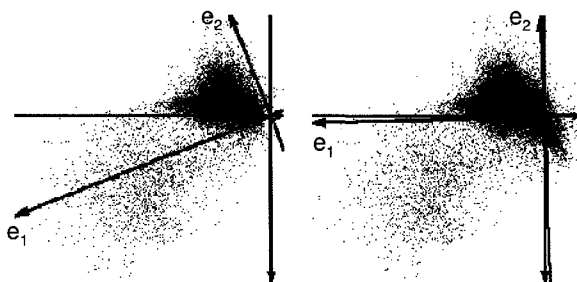


Fig. 8. Original axis and PCA directions in the feature space for the bimodal ( $228 \times 228$ ) data set (left) and for the trimodal ( $414 \times 326$ ) data set (right). The dark cluster close to the origin corresponds to the "no-change" class, whereas the light-gray one corresponds to the "change" class.

Similar results to those of the K&I and Fisher/EM methods were also yielded by Bruzzone and Prieto's method on the 3-mode data set. The performances of their algorithm were slightly worse than those of the Fisher/EM method and of K&I with PCA-1, in the case of the 2-mode data set. On the considered data set, Bruzzone and Prieto's algorithm proved to be insensitive to the choice of the initial thresholds,  $T_c$  and  $T_n$  (see Section 2).

## 6. Conclusions

In this chapter, two unsupervised change-detection methodologies, operating in the context of image differencing, have been described and compared with another reference algorithm. Two unsupervised transforms, namely, the first component of PCA (PCA-1) and the modulus, have been used, in a preliminary step, to transform the difference image and obtain scalar images.

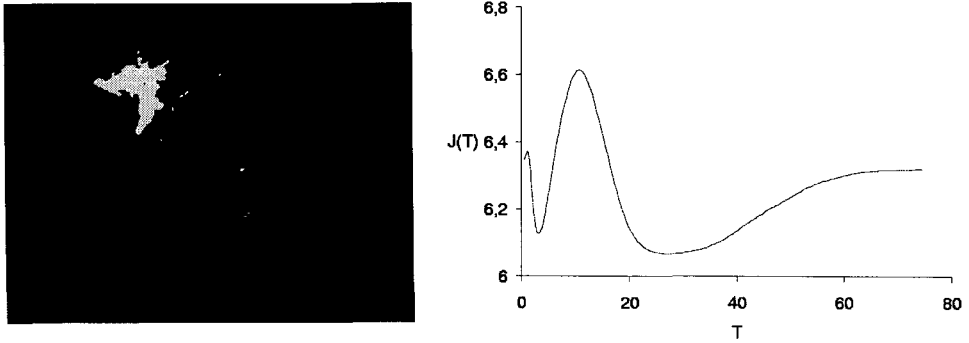


Fig. 9. Results of the K&I algorithm applied to the modulus image of the  $414 \times 326$  data set: change map (left) and criterion function (right).

A difference between the PCA-1 and modulus operators, both applied to an input Gaussian mixture, lies in the fact that the PCA-1 is linear, whereas the modulus is not. Therefore, PCA-1 images save the Gaussian property, whereas the modulus image does not maintain normal statistics. The consequences are shown by the experimental results on the reduced data set (which exhibits a 2-mode distribution): the thresholding K&I algorithm performed better when applied to PCA-1 than when applied to the modulus. However, in the experiments with the complete data set (which exhibits a 3-mode distribution), K&I seemed to suffer from the trimodality more when applied to PCA-1 than when applied to the modulus image. This can be explained considering that the presence of a third mode may reduce the capability of PCA-1 to maintain the separation between the "change" and "no-change" classes in its first component. From a general viewpoint, PCA-1 is an image-dependent operator, and is consequently influenced by the data statistics. On the other hand, the modulus operator is effective in separating the classes only under an important assumption, i.e., low modulus values should correspond to the "no-change" class and high values to the "change" one. This is usually true for change detection in remote sensing. However, especially when there are changes in radiance also for the "no-change" class, this assumption should be properly checked.

The considered thresholding method, i.e., K&I, provided very good results with the two-mode reduced data set. Moreover, it is a very fast approach, as no iterations are needed but only the calculation of the criterion function, which is defined for  $L$  values (e.g., 256 values). In addition, given the histogram, the computation time is independent of the image size, for it depends only on the number of grey levels in the input image.

The Expectation-Maximization algorithm proved to be an effective approach to parameter estimation, robust and highly insensitive to the initial parameter vector. Therefore, it appears to be an appropriate estimation tool on which the classification step may rely (at least for the problem faced in this paper). In particular, the

EM modified version<sup>42</sup> exhibited overall better performances in all the experiments, and resulted in more precise parameter estimates. Moreover, it always reached convergence faster than the standard version. EM iterations use only grey-level information, so they can be implemented in such a way that, given the input image histogram, the computation time turns out to be independent of the image size.

Concerning the method based on the Fisher transform, utilized in conjunction with the EM algorithm, the achieved results were slightly better than those yielded by K&I after the unsupervised transform. Furthermore, the Fisher/EM methodology exhibited a slightly lower sensitivity to the inaccuracy of the Gaussian model for the class-conditional distribution as compared with K&I using PCA-1. Fisher/EM and K&I showed a stable and regular behavior in the presence of additive Gaussian noise in the image.

Comparing Fisher/EM with Bruzzone and Prieto's method, the slightly better results obtained by the former method suggest that the use of the Fisher transform can improve the discrimination between the two classes "change" and "no-change" thus leading to a more accurate thresholding. From a computational viewpoint, the complexity of Fisher/EM is slightly higher than that of Bruzzone and Prieto's method, whereas K&I was by far the least time-consuming technique.

To sum at, the K&I technique, originally developed for artificial-vision problems, turned out to be a valuable tool for change detection in remote-sensing images, as, in spite of its simplicity, it provided very accurate change maps. The results reported in this chapter confirm that improved results can be obtained by Bruzzone and Prieto's method, adopting the Jackson-Landgrebe version of EM. Finally, the proposed combination of feature transformation by the Fisher transform with parameter estimation by EM appears to be a powerful approach able to provide the best change-detection results among the considered techniques.

## Acknowledgments

This research was funded in part by the Italian Ministry of Education, University and Research. The support is gratefully acknowledged.

## References

1. R. A. Washington-Allen, R. D. Ramsey, B. E. Norton, and N. E. West, "Change detection of the effect of severe drought on subsistence agropastoral communities on the Bolivian Altiplano," *International Journal of Remote Sensing*, vol. **19**, no. 7, pp. 1319-1333, 1998.
2. A. Y. Kwarteng, and P. S. Chavez, "Change detection study of Kuwait City and environs using multi-temporal Landsat Thematic Mapper data," *International Journal of Remote Sensing*, vol. **19**, no. 9, pp. 1651-1662, 1998.
3. E. M. Pereira, and A. W. Setzer, "Spectral characteristics of fire scars in Landsat-5 TM images of Amazonia," *International Journal of Remote Sensing*, vol. **14**, no. 11, 1993.
4. R. F. Nelson, "Detecting forest canopy change due to insect activity using Land-

- sat MSS," *Photogrammetric Engineering & Remote Sensing*, vol. **49**, pp. 1303-1314, 1983.
5. J. R. G. Townshend, C. O. Justice, and C. Gurney, "The impact of misregistration on change detection," *IEEE Trans. on Geoscience and Remote Sensing*, vol. **30**, no. 5, pp. 1054-1060, 1992.
  6. P. Gong, E. F. Ledrew, and J. R. Miller, "Registration-noise reduction in difference images for change detection," *Int. Journal of Remote Sensing*, vol. **13**, no. 4, pp. 773-779, 1992.
  7. P. N. Slater, "Reflectance and radiance based methods for the in-flight absolute calibration of multispectral sensors," *Remote Sensing of Environment*, vol. **22**, pp.11-37, 1987.
  8. P. S. Chavez, "Radiometric calibration of Landsat Thematic Mapper multispectral images," *Photogrammetric Engineering & Remote Sensing*, vol. **55**, no. 9, pp. 1285-1294, 1989.
  9. P. S. Chavez and D. J. MacKinnon, "Automatic detection of vegetation changes in the southwestern United States using remotely sensed images," *Photogrammetric Engineering & Remote Sensing*, vol. **60**, no. 5, pp. 1285-1294, 1994.
  10. A. Singh, "Digital change detection techniques using remotely-sensed data," *International Journal of Remote Sensing*, vol. **10**, no. 6, pp. 989-1003, 1989.
  11. H. L. Van Trees, *Detection, Estimation and Modulation Theory*, Part I, New York: Wiley, 2001.
  12. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd Edition. New York: Wiley, 2001, pp. 117-121.
  13. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd Edition. New York: Academic Press, 1990, pp. 26-34.
  14. L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection", *IEEE Trans. on Geoscience and Remote Sensing*, vol. **38**, pp. 1171-1182, 2000.
  15. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *J. Royal Statist. Soc.*, vol. **19**, n. 1, pp. 1-38, 1977.
  16. T. K. Moon, "The expectation-maximization algorithm", *Signal Processing Mag.*, vol. **13**, n. 6, pp. 47-60, 1996.
  17. T. Aach and A. Kaup, "Bayesian algorithms for adaptive change detection in image sequences using Markov random fields", *Signal Processing: Image Communication*, vol. **7**, pp. 147-160, 1995.
  18. L. Bruzzone, and D. F. Prieto, "An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images", *IEEE Trans. Image Processing*, vol. **11**, pp. 452-466, 2002.
  19. T. Kasetkasem, and P. K. Varshney, "An image change detection algorithm based on Markov Random Field models", *IEEE Trans. Geosci. Remote Sensing*, vol. **40**, pp. 1815-1823, 2002.
  20. F. Melgani, G. Moser, S. B. Serpico, "Unsupervised change detection methods for remote sensing images", *Optical Engineering*, in press, 2002.
  21. P. H. Swain, "Bayesian classification in a time-varying environment", *IEEE Trans. Syst., Man, Cybern.*, vol. **8**, pp. 879-883, 1978.
  22. H. M. Kalayeh and D. A. Landgrebe, "Utilizing multitemporal data by a stochastic model", *IEEE Trans. Geosci. Remote Sensing*, vol. **24**, pp. 792-795, 1986.
  23. L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images", *IEEE Trans. Geosci. Remote Sensing*, vol. **35**, pp. 858-867, 1997.

24. L. Bruzzone, D. F. Prieto, and S. B. Serpico, "A neural-statistical approach to multi-temporal and multisource remote-sensing image classification", *IEEE Trans. Geosci. Remote Sensing*, vol. **37**, pp. 1350-1359, 1999.
25. B. Jeon and D. A. Landgrebe, "Decision fusion approach for multitemporal classification", *IEEE Trans. Geosci. Remote Sensing*, vol. **37**, pp. 1227-1233, 1999.
26. B. Tian, M. R. Azimi-Sadjadi, T. H. Vonder Haar, and D. Reinke, "Temporal updating scheme for probabilistic neural network with application to satellite cloud classification", *IEEE Trans. Neural Networks*, vol. **11**, pp. 903-920, 2000.
27. N. Khazenie, M. M. Crawford, "Spatio-temporal autocorrelated model for contextual classification", *IEEE Trans. Geosci. Remote Sensing*, vol. **28**, pp. 529-539, 1990.
28. B. Jeon and D. A. Landgrebe, "Classification with spatio-temporal interpixel class dependency contexts", *IEEE Trans. Geosci. Remote Sensing*, vol. **30**, pp. 663-672, 1992.
29. A. H. S. Solberg, T. Taxt, and A. K. Jain, "A Markov Random Field model for classification of multisource satellite imagery", *IEEE Trans. Geosci. Remote Sensing*, vol. **43**, pp. 100-113, 1996.
30. F. Melgani and S. B. Serpico, "A statistical approach to the fusion of the spectral and spatio-temporal contextual information for the classification of remote sensing images", *Pattern Recognition Letters*, vol. **23**, no. 9, pp. 1053-1061, April 2002.
31. L. Bruzzone and D. F. Prieto, "Unsupervised retraining of a maximum-likelihood classifier for the analysis of multitemporal remote-sensing images", *IEEE Transactions on Geoscience and Remote Sensing*, vol. **39**, pp. 456-460, 2001.
32. D. F. Prieto and O. Arino, "A partially supervised change detection technique", *Proc. of IEEE-IGARSS 2001*, Sydney, vol. **2**, pp. 858-860, 2001.
33. G. Moser, F. Melgani, S. B. Serpico, and A. Caruso, "Partially Supervised Detection of Changes from Remote Sensing Images", *Proc. of IEEE-IGARSS 2002*, Toronto, vol. I, pp. 299-301.
34. Z. Chi, H. Yan, T. Pham, *Fuzzy algorithms: with applications to image processing and pattern recognition*, Singapore: World Scientific Publishing, 1996, pp. 45-84.
35. J. Kittler and J. Illingworth, "Minimum error thresholding", *Pattern Recognition*, vol. **19**, pp. 41-47, 1986.
36. L. K. Huang and M. J. Wang, "Image thresholding by minimizing the measures of fuzziness", *Pattern Recognition*, vol. **28**, pp. 41-51, 1995.
37. N. Otsu, "A threshold selection method from gray-level histogram", *IEEE Trans. on Syst., Man., Cybern.*, vol. **SMC-9**, pp. 62-66, 1979.
38. G. Moser, "Development of unsupervised change detection methods for remote sensing images", "Laurea" degree thesis, University of Genoa, November 2001.
39. A. Papoulis, Probability, *Random Variables and Stochastic Processes*, New York: McGraw-Hill Book Company, 2nd edition, 1984.
40. Y. Delignon, A. Marzouki, W. Pieczynski, "Estimation of generalized mixtures and its application in image segmentation", *IEEE Trans. Image Processing*, vol. **6**, pp. 1364-1374, 1997.
41. C. F. Jeff Wu, "On the convergence properties of the EM algorithm", *The Annals of Statistics*, vol. **11**, pp. 95-103, 1983.
42. Q. Jackson, D. Landgrebe, "An adaptive classifier design for high-dimensional data analysis with a limited training data set", *IEEE Trans. Geosci. Remote Sensing*, vol. **39**, pp. 2664-2679, 2001.
43. X. Yang, C. P. Lo, "Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area", *Int. J. Remote Sensing*, vol. **23**, no. 9, pp. 1775-1798, 2002.



44. L. Jimenez, and D. Landgrebe, "Supervised Classification in High Dimensional Space: Geometrical, Statistical, and Asymptotical Properties of Multivariate Data," *IEEE Transactions on System, Man, and Cybernetics*, vol. **28**, Part C, No. 1, pp. 39-54, February 1998.

# **Seismic Signal Processing**

This page is intentionally left blank

# CHAPTER 19

## REMOTE DETECTION USING DUAL SENSORS

Enders A. Robinson

*Columbia University, Krumb School of Mines, 500 West 120 Street  
New York, NY 10027 USA*

The reflection seismic method is an instrument for remote detection that uses traveling waves to find the structure of an inaccessible body. The model used consists of flat horizontally layers subjected to seismic compressional waves at normal incidence. Both the source and receiver are buried below the surface with the receiver below the source. Dual attributes of the seismic wavefield are particle velocity and pressure. A receiver to measure these dual attributes is made up of two sensors. One sensor is a geophone, which measures particle velocity. The other sensor is a hydrophone, which measures pressure. Einstein deconvolution makes use of these dual attributes. The method is called Einstein deconvolution because the mathematics involved is similar to the mathematics of the special theory of relativity. Einstein deconvolution consists of two steps. The first step is the use of the d'Alembert equations to convert the received particle-velocity signal and the received pressure signal into the downgoing and upgoing waves. This step requires knowledge of the acoustic impedance of the material at the receiver. The second step in the Einstein deconvolution process is to deconvolve the upgoing wave by the downgoing wave. This operation removes the unknown source signature (which may or may not be minimum-phase) as well as the reverberations and ghosts due to the layers above the receiver. The output of the Einstein deconvolution process is the unit-impulse reflection response of the layers below the receiver. This unit-impulse reflection response can be subjected to dynamic deconvolution to yield the individual reflection coefficients.

### 1 Introduction

The reflection seismic method considered as an instrument for remote detection has much in common with other disciplines that use non-invasive techniques to find the structure of an inaccessible body. Other such disciplines are medical imaging and non-destructive evaluation. Generally some sort of wall or barrier protects the remote body. In the seismic case the wall is made up of the surface layers of the earth; in the medical case the wall is the human skin. A source of energy must be chosen to penetrate the wall and to reach the structure beyond. Receivers are placed to pick up the information-bearing signals. Often the source and receivers are restricted to lie outside of the wall; sometimes they can be placed a certain distance inside the wall. For example, in the case of an echocardiogram, the source and receivers are placed on the surface of the skin. In certain laser applications the source and receiver are on an instrument that penetrates the skin. In vibroseis seismic prospecting on land, the source and receivers are placed on the surface of the earth. In seismic prospecting at sea, the source and receivers are placed below the surface of the water.

The work of the geophysicist is to look beneath the surface of the earth in the search for new deposits of oil and natural gas. The subsurface geologic structures of interest can be as deep as four or five miles. The exploration geophysicist illuminates the subsurface by means of an energy source that generates seismic waves. The subsurface rock layers transmit and reflect the seismic waves. A seismic survey consists of collecting data over a selected geographic area, called the prospect. Essential features of seismic data acquisition are: At a fixed point on or near the surface of the earth, a source of energy, such as arrays of dynamite charges, or air guns, or chirp-signal (vibroseis) generators, is activated. An activated source is called a shot. Seismic waves from the shot propagate downward from the source point deep into the earth. The waves are reflected from geologic interfaces. The reflected waves propagate upward from the reflecting interfaces. A primary reflection has a wavepath that goes directly down to the reflecting interface, and then directly back up to the surface. A multiple reflection (or reverberation) has a wavepath that goes back and forth among various interfaces as the wave proceeds on its trip. Receivers at or near the surface detect the reflected waves, both primaries and multiples. If the source is buried below the surface of the earth, the surface layers reflect the upward traveling blast of the source energy, thereby producing a secondary blast of downgoing energy in addition to the direct blast of downgoing energy from the source. This secondary blast produces what are known as ghost reflections. The digitized signal recorded at a receiver is called a trace. After one shot is completed, the source point and the corresponding receiver points are moved so that another shot can take place. This acquisition method is repeated again and again until the entire prospect is covered.

The traces as recorded by the receivers constitute the raw data, which are then fed into computers for processing. The purpose of digital seismic processing is to transform the raw data into computer-generated images of the subsurface geological structures. The high dynamic range of the receivers available today makes possible the use of precise signal processing methods that give excellent image quality. Geologists and geophysicists then interpret the images in the form of maps and cross-sections in order to choose the most favorable drilling sites for new oil wells, either wildcats or field-extension wells.

Geophysicists have at their disposal the most advanced computer power available. By using commercially available seismic processing packages, a geophysicist can process seismic data even if he knows little or no mathematics at all. The results obtained are of a quality that would have been beyond the reach of the most skilled mathematician in the days before the computer. Consequently, a geophysicist now can dispense with much of the complicated type of mathematics that is needed to describe and analyze seismic wave propagation. What can be said about the value of mathematics in geophysics today? A geophysicist with a practical knowledge of mathematics is more able to solve those difficult problems that inevitably arise from time to time in seismic processing. A geophysicist can use mathematics to sort out the effects of the various processing operations. In this way, mathematics gives a geophysicist something that is quite difficult to obtain through the use of commercially available software alone. That illusive something is insight.

A word should be said about velocity, as the term is used in two different contexts in geophysics. Particle velocity refers to the velocity of a small particle of rock, as the particle moves back and forth about its equilibrium position due to the passage of a wave. Propagation velocity (or wave velocity) refers to the velocity of a traveling wave as it propagates from one place to some other place.

In this paper the Z-transform is the main mathematical vehicle used. The Z-transform as used in geophysics is, in fact, the generating function as used in mathematics and statistics. To convert a geophysical Z-transform to the corresponding electrical engineering Z-transform, replace each  $Z$  in the expression by  $z^{-1}$ . We represent waves by their respective Z-transforms. Much can be learned about geophysical signal processing through the relatively simple mathematics of the Z-transform. The operations of convolution and deconvolution for signals become respectively the multiplication and division of their Z-transforms. Robinson (1966) gives two geophysical processing methods, which he describes in terms of the mathematics of the Z-transform. One is the method for the elimination of seismic ghost reflections, and the other is the method of predictive deconvolution for the elimination of water-confined reverberations. These two processing methods are still in use today in the same form as originally given. The reason is that the robustness of the two methods makes their application much wider than otherwise would be expected.

The purpose of the present paper is to meld the above two processing methods into one overall deconvolution method that makes use of the dual attributes of particle velocity and pressure. This overall deconvolution method is called Einstein deconvolution (Robinson, 1999). The method is named after Albert Einstein because the mathematics involved is similar to the mathematics used in the special theory of relativity. Einstein deconvolution eliminates the near-surface reverberations, the ghost reflections, and the unknown source signature. We treat the case where both the source and the receiver are placed below the surface. In the case of land exploration, the surface is the surface of the ground. In the case of marine exploration, the surface is the surface of the water. Einstein deconvolution removes all the extraneous effects above the receiver. For this reason, the receiver should be placed below the source, so that all the extraneous effects due to the source are removed as well. In this paper we use a model based upon compressional waves traveling at normal incidence to flat horizontal interfaces.

The complete solution of the wave equation requires that two boundary conditions be satisfied. Measurement of the dual attributes of particle-velocity and pressure by a dual sensor can satisfy that requirement. According to the classical result of d'Alembert, the general solution of the wave equation in any rock layer yields wave motion that is the sum of two components. One component is the downgoing wave and the other component is the upgoing wave. In the first step, Einstein deconvolution converts the recorded particle-velocity signal and the recorded pressure signal (as measured by the dual sensor) into the downgoing traveling wave and the upgoing traveling wave. In order to carry out this step the acoustic impedance of the rock layer at the receiver must be

known or estimated. In the second step, Einstein deconvolution uses the downgoing wave to deconvolve the upgoing wave. This step removes the unknown source signature, which may be minimum-phase or may not be. (The term minimum-delay, often used in geophysics, means the same thing as minimum-phase.) This step also removes the reverberations and ghosts due to the layers above the receiver. Since this type of noise is the most harmful on a marine seismogram, Einstein deconvolution has widespread applications in exploration. In fact, the output of Einstein deconvolution is the unit-impulse reflection response of the layers below the receiver. If desired, this unit-impulse reflection response can be subjected to dynamic deconvolution to yield the individual reflection coefficients.

## 2 Background on the Use of Dual Sensors

Einstein deconvolution requires the acquisition of seismic data making use of dual sensors. The main use of dual sensors today is in connection with ocean bottom cable (OBC) work. Haggerty (1956) did fundamental work on dual sensors for exploration at sea. A typical dual sensor is made up of a geophone and a hydrophone. Such a dual sensor simultaneously records two signals. The geophone measures the particle velocity, whereas the hydrophone measures the pressure. Haggerty also describes a marine seismic reflection surveying system in which two seismometer spreads are disposed at two distinct depths in the water such that water column reverberations received by them are 180 degrees out of phase. By combining the output of the detectors, the reverberations cancel. Silverman et al (1963) describe a multiple-reflection analog computer, which produces all primary and multiple reflections, and which allows pressure and particle-velocity detectors to be placed at any depth. Schneider and Backus (1964) describe the use of a dual sensor on the ocean floor as part of the Vela Uniform program to measure and study ocean-bottom seismic phenomenon in the frequency range of 0.5 Hz to 10 Hz. White (1965) describes a dual sensor located near a water-solid interface and considers a fluid-borne wave as well as a compressional wave arriving at normal incidence from the solid. White then derives the response when the geophone trace is multiplied by a scale factor and the product is added to the hydrophone trace. Robinson (1966) introduces predictive deconvolution for the removal of water-column reverberations with the requirement that a transducer that measures both the downgoing wave-motion and the upgoing wave-motion should be utilized. In the usual embodiment of predictive deconvolution the upgoing wave is deconvolved. Robinson (1984, p. 113) derives the d'Alembert equations that give the downgoing traveling motion and upgoing traveling motion in terms of particle-velocity, pressure, and the acoustic impedance at the receiver.

Pavey (1966) describes the placement of dual sensors in a towed marine streamer for the elimination of ghost reflections. Furthermore Pavey (1967) provides a method for canceling the ghost wave reflected downward from the surface of the water by the use of particle-velocity sensing detectors interspersed with pressure sensing detectors within the streamer. Gal'perin (1971) in his work on vertical seismic profiling considers the use of a

dual sensor in a borehole to simplify the wavefield observed at interior points of the medium. Loewenthal (1975), as a general method to remove unwanted signals, put forth the first systematic use of dual data in connection with the magnetotelluric inverse problem. Robinson (1975) shows that the Einstein addition formula (Einstein, 1905) has the same mathematical structure as the equation for the unit-impulse reflection response obtained when an additional layer is added to the medium. Cowles (1979) describes a method and apparatus for measuring and identifying up-traveling and down-traveling waves. Both a particle-velocity sensitive transducer and pressure sensitive transducer are placed at a point below the earth's surface, as for example, in a borehole that is refilled. Ziolkowski et al. (1980) present a method for the extraction of the source wavelet from the reflection seismogram. They generate two different seismograms from each source-receiver pair where the source used to generate one seismogram is a scaled version of the source used to generate the other.

Ruehle (1984) gives a technique for reducing ghosting in which a pressure detector and a particle-velocity detector are positioned in close proximity to one another in the water. The pressure detector produces a positive output in response to upwardly traveling compressional waves and a negative output in response to downwardly traveling rarefaction waves. The particle-velocity detector produces a positive output in response to upwardly traveling compressional waves and a positive output in response to downwardly traveling rarefaction waves. These outputs are filtered so that the impulse response of the rarefaction waves cancels. The filtered outputs are combined to produce an output in which the ghost reflection is substantially suppressed. Berni (1984) describes the use of a vertical component accelerometer in combination with a hydrophone for canceling surface-reflected ghosts in marine seismic operations. Berni (1985) gives a method for eliminating ghosts from seismic signals detected at a predetermined depth. The method employs both a pressure sensor and a motion sensor. A seismic signal and its corresponding ghost signal detected by a pressure sensor is filtered as a function of depth of the sensor to provide a band-limited spike signal at an arrival time midway between the seismic and ghost arrival times. Similarly, a seismic signal and its corresponding ghost signal detected by a motion sensor is filtered as a function of depth of the sensor to provide a band-limited spike signal at an arrival time midway between the seismic and ghost arrival times. The filtered motion and pressure spikes are then added together in proportion to their respective signal-to-noise ratios. The proportionally added signal is a ghost free seismic signal having a maximum signal-to-noise ratio for each frequency component and may be employed in further seismic processing steps. Szaraniec (1985) uses the odd-depth layered-earth model in the Z-transform domain directly to find the impulse response. Gutowski and Treitel (1987) construct a normal-incidence synthetic seismogram in the case of arbitrary source and receiver positions. Adair et al. (1988) compare signal-to-noise ratios obtained on bottomed seismometers, bottomed hydrophones, and buried seismometers from near-surface explosions. Using the concept of a fictitious surface source, Shtivelman and Loewenthal (1989) give an alternative and more intuitive process of construction of a



normal-incidence synthetic seismogram. Their final results are similar to those of Gutowski and Treitel (1987) but are obtained quite differently.

Loewenthal and Jakubowicz (1983) and Loewenthal et al. (1985) demonstrate a deterministic approach for estimating the wavelet and reflectivity for one-dimensional seismograms. The technique enables the signature of any point source or, in principle, any horizontal source array to be determined for a one-dimensional earth. The resulting signature is of the far-field form and includes the source ghost. Provided the source wavelet is space-invariant, that wavelet is convolved in both the recorded pressure and particle-velocity fields. It follows that the ratio of the Fourier transform of particle velocity to that of pressure is always source independent. Due to the cancellation of the source term, the method always yields an impulse response. This can be equivalently viewed as an extrapolation of the wavefields from receivers at depth to receivers placed at the surface. Barr and Sanders (1989) give a method to remove multiples through the combination of pressure and particle-velocity measurements. They show that the dual-sensor method eliminates notches in the recorded OBC data's amplitude spectra that would otherwise be generated by the receiver's location below the water's surface. Towed streamer data, however, continue to contain such notches, resulting in reduced bandwidth. The dual-sensor method records both particle-velocity and pressure signals in ocean-bottom cable surveys. Elimination of surface reflected energy is obtained by simple summation of the pressure and velocity signals. Barr and Sanders (1989) and Barr (1990) show that if a suitably scaled particle-velocity trace is used in the summation, both the upward and the downward traveling parts of the wavefield trapped in the water layer can be eliminated. Their scaling factor is a function of the ocean-bottom reflectivity. Early dual-sensor surveys used a separate calibration survey to measure the ocean-bottom reflectivity. In the calibration survey a low-energy source was fired over each receiver pair and the scalar was taken as the ratio of the first break amplitudes of the hydrophone and geophone signals. The added cost of such a survey led to the development of methods to estimate the reflectivity from the production data itself.

Barr (1990) describes a marine seismic system that reduces coherent noise by applying a scale factor to the output of a pressure sensor and a particle-velocity sensor positioned adjacent to one another in the water. The sensors can be placed at a point in the water above the bottom in order to eliminate downgoing components of the reverberation. Alternatively, they can be placed on the water's bottom in order to eliminate both upgoing and downgoing components of the reverberation. The scale factor, which depends upon the acoustical impedance of the water or water-bottom material, can be determined either deterministically or statistically. The deterministic method involves measuring and comparing the responses of the pressure and particle-velocity sensors to a pressure wave induced in the water. The statistical method involves comparing the magnitude of the pressure-signal autocorrelation to the crosscorrelation of the pressure-signal and the particle-velocity-signal at selected lag values or, alternatively, comparing the magnitude of the pressure-signal autocorrelation to the particle-velocity signal autocorrelation at selected lag values. Paffenholz and Barr (1995) point out that

one of the primary goals of the dual-sensor technology is the elimination of the receiver ghost response that corrupts single hydrophone ocean-bottom cable data. The dual-sensor data allow separation of the wavefield into upward and downward traveling components. They show how the elimination of the surface reflected energy can be achieved by summation of the pressure and particle-velocity signals. They show that this dual-sensor processing requires the use of scale factors that are functions of the ocean-bottom reflectivity. They give a new method to derive the ocean-bottom reflectivity from production seismic data. Their method eliminates the need for a separate calibration survey and is stable in the presence of random and shot-generated noise. They also discuss methods for determining the receiver depths from production seismic data.

Barr, Paffenholz, and Rabson (1996) make important contributions to the ocean-bottom cable (OBC) method. They discuss the advantages of not using streamers, but placing the detectors on the ocean bottom. They show that the method can safely deal with obstacles like oil-field equipment platforms. In addition, they show that OBC data, enhanced with numerous data acquisition and processing technologies, have a resolution that surpasses the resolution achievable with today's towed streamer technology. Because the OBC method employs a stationary array of receiver stations on the ocean bottom and a marine vessel towing only a seismic energy source, they show that the physical separation of the energy source from the recording spread provides the flexibility to record virtually any geometry. Also they show that the OBC method's stationary receiver spread also yields a surface consistent recording geometry. Sun (1997) describes the vertical cable method for acquiring and processing prestack 3-D marine seismic data. The method is based on technology developed by the US Navy for antisubmarine warfare. Processing schemes used for the separation of upgoing and downgoing wavefields of vertical cable data are quite different from the schemes for separation of the upgoing and downgoing wave fields in standard VSP data processing. Sun (1997) develops a new method for the separation of upgoing and downgoing wave fields. The method achieves the maximum utilization of vertical cable field data. In tests with synthetic data and field vertical cable data, the method not only performs well but it is computationally simple since it assumes a stratified earth model and requires only the water-layer velocity. Other papers on dual-field techniques include Shtivelman and Loewenthal (1988), Loewenthal and Shtivelman (1989), and Loewenthal and Stoffa (1991), Loewenthal (1991, 1994), Barr (1997), and Dragoset and Barr (1994). Canales and Bell (1996) combine hydrophone and geophone data to separate the upgoing and downgoing waves. Then they use the downgoing wave as the effective source wavelet and minimize the energy of the upgoing wavefield deconvolved by the downgoing wavefield. Their method improves continuity and attenuates the ghost.

The Einstein deconvolution method described in the paper is limited to compressional waves, flat interfaces, and normal incidence. Standard techniques such as the Radon transformation can be used for the case of non-normal incidence. Aminzadeh (1984) develops a recursive algorithm to obtain the layer parameters of an elastic medium (density, P-wave velocity, S-wave velocity) from reflection coefficient matrices in terms

of energy flux ratios for a non-normal incidence case. Recent work is moving the technology to non-normal incidence approximations and includes correct scaling of data through knowledge of the sea-floor parameters. Recent works also addresses four-component data recording. The principal value of OBC data is the recording of converted waves to help characterize the sub-surface. A generalization is required in order to make the present methods applicable to converted-wave multi-component OBC data.

### 3 The Source Signal

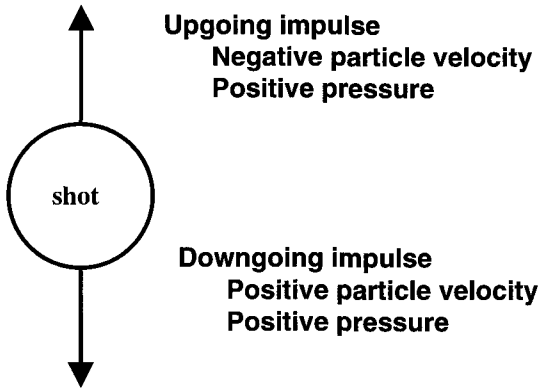


Fig. 1. The convention used.

Let us now consider an explosion in a shot hole buried below the surface of the ground or, alternatively, an air-gun discharge below the surface of the water. See Fig. 1. For the moment assume that the shot produces an upgoing spike and a downgoing spike. The positive direction of the vertical axis is downward, and the negative direction of the vertical axis is upward. Let the downgoing spike in particle velocity be denoted by  $\delta^d$ . Particle velocity has a direction associated with it, so the upgoing spike  $\delta^u$  in particle velocity has the opposite sign to the downgoing spike in particle velocity; that is  $\delta^u = -\delta^d$ . On the other hand, pressure is not a directional quantity, so the upgoing spike in pressure has the same sign as the downgoing spike in pressure. In summary, according to our convention, an ideal shot has this characteristic: the downgoing spike (which travels in the positive vertical direction) has positive particle velocity and positive pressure, and the upgoing spike (which travels in the negative vertical direction) has negative particle velocity but still a positive pressure. In actuality a physical seismic source produces not a spike, but a signal  $S$  of some duration called the signature. As a result the downgoing source signature in particle velocity produced by a realistic source can be written as  $\dot{S}$ . The upgoing source signature in particle velocity can be written as

–S. For Einstein deconvolution, the signature is not required to be minimum-phase. In fact, for Einstein deconvolution, the signature can be a completely unknown signal.

#### 4 Fresnel Reflection and Transmission Coefficients

A seismic wave involves both pressure and particle velocity. Augustin Jean Fresnel (1788–1827) provided the basic concepts for the definition of a reflection coefficient. In seismic work, the reflection coefficient of an interface is found by requiring that (a) the particle velocity be continuous across the interface and (b) the pressure be continuous across the interface. In this paper we make use of the classical one-dimensional homogeneous isotropic parallel-layered-earth model (Robinson, 1967).

The layered-earth model can be described as follows. Layer  $i$  is defined as the layer between upper interface  $i-1$  and lower interface  $i$ . Denote the density, or mass per unit volume, of layer  $i$  by  $\rho_i$ . Let the propagation velocity of the wave motion in that layer be  $C_i$ . The product  $\rho_i C_i$  is called the acoustic impedance of the layer. In that layer, denote the solution of the wave equation for particle velocity by  $V_i$  (which we call the particle-velocity signal), and denote the solution of the wave equation for pressure by  $p_i$  (which we call the pressure signal). These two signals represent the dual field.

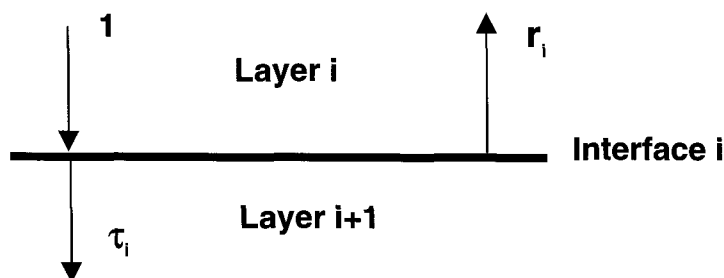


Fig. 2. The Fresnel reflection coefficient  $r_i$  and the transmission coefficient  $\tau_i$ .

Consider a single interface, say interface  $i$ , in isolation. See Fig. 2. When a downgoing wave strikes the interface, part of the energy is reflected back into the same layer and the rest of the energy is transmitted into the next layer. For an incident downgoing particle-velocity signal striking the interface from above, the particle-velocity Fresnel reflection coefficient  $r_i$  and the particle-velocity Fresnel transmission coefficient  $\tau_i$  are given by

$$r_i = \frac{\rho_i C_i - \rho_{i+1} C_{i+1}}{\rho_i C_i + \rho_{i+1} C_{i+1}} \quad ; \quad \tau_i = 1 + r_i = \frac{2 \rho_i C_i}{\rho_i C_i + \rho_{i+1} C_{i+1}} \quad (1)$$

These reflection coefficients are called Fresnel coefficients because they deal with only the interface between the two layers in question, and not with the entire system

which can have many layers. Each Fresnel coefficient carries the number of the interface as a subscript. It is seen that the reflection coefficient must be less than one in magnitude. Consider next an incident upgoing particle-velocity signal in layer  $i + 1$  striking the interface from below. The resulting Fresnel reflection and transmission coefficients carry a prime, and are given by

$$r'_i = -r_i \quad ; \quad \tau'_i = 1 + r'_i = 1 - r_i = \frac{\rho_{i+1} C_{i+1}}{\rho_i C_i} \tau_i \quad (2)$$

## 5 Elimination of Seismic Ghost Reflections

Robinson (1966) derives a filter for the elimination of ghost reflections. In seismic exploration the explosion is set off in a shot hole drilled below the surface of the ground. The existence of a large velocity discontinuity above the seismic shot can act as the source of ghost reflections in the following way. The primary reflections on the seismogram are caused by the reflection from deep strata of the energy moving directly downward from the shotpoint. Meanwhile the energy moving directly upward from the shotpoint is reflected from the overlying discontinuity, and thus there is a source of secondary energy moving directly downwards. This secondary energy in turn is reflected from the deep strata, thereby producing corresponding ghost reflections on the seismogram. Thus a deep reflecting horizon appears on the seismogram as two reflection wavelets, displaced in time by twice the traveltime from the shot to the overlying discontinuity. Any differences in shape between the primary and ghost reflections can be attributed to various causes. However, in many instances the primary and ghost have approximately the same shape. In such a case, the two most important parameters become  $r'_A$  and  $n$ . The constant  $r'_A$  (of magnitude less than unity) is the Fresnel reflection coefficient of the overlying interface A subject to an upgoing incident wave, and the constant  $n$  (assumed to be an integer) representing the time-delay of the ghost with respect to the primary.

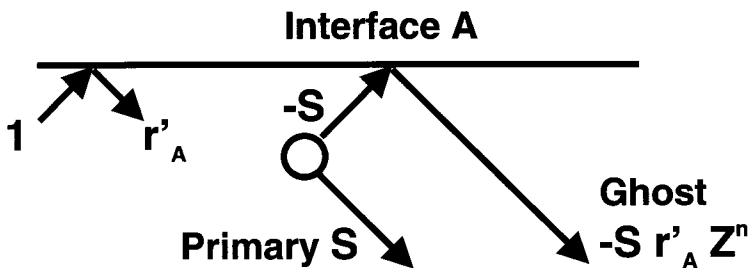


Fig. 3. Ghost-producing couplet.

The downgoing source signature in particle velocity is  $S$  and the upgoing source signature in particle velocity is  $-S$ . See Fig. 3. Hence the ghost-producing couplet (in particle velocity) may be represented as

$$S - S r'_A Z^n = S(1 - r'_A Z^n) \quad (3)$$

It follows that the ghost-producing filter is the couplet

$$G = 1 - r'_A Z^n \quad (4)$$

This couplet is minimum-phase because its initial coefficient is greater in magnitude than its final coefficient, so more of the energy is concentrated at the front than at the end. A filter  $F$  to eliminate ghost reflections would be one that converts a primary-ghost couplet at the input into a primary alone at the output. The ghost elimination filter is the inverse of the couplet (4) as given by

$$F = \frac{1}{1 - r'_A Z^n} \quad (5)$$

Because of the linear property of the filter, a train of overlapping primary reflections plus their ghosts is transformed by the filter into a train of primary reflections only. There are various ways of estimating the required parameters from the seismic data for the design of the prototype ghost elimination filter described above. Also more refined designs of ghost elimination filters may be used. For example, a ghost elimination filter may be designed as the least-squares shaping filter whose input is the waveform made up of the primary with its ghost, and the desired output is the waveform of the primary only.

## 6 Elimination of Water Reverberations

Robinson (1966) gives the method of predictive deconvolution for the elimination of water reverberations. The water reverberation problem in marine seismic operations may be described as follows. The water-air interface is a strong reflector. Let the water-bottom interface also be a strong reflector. In such a case the water layer approximates a medium bounded by two strong reflecting interfaces and hence represents an energy trap. A seismic pulse generated in this energy trap will be successively reflected between the two interfaces. Consequently, the water reverberations will obscure reflections from deep horizons below the water layer.

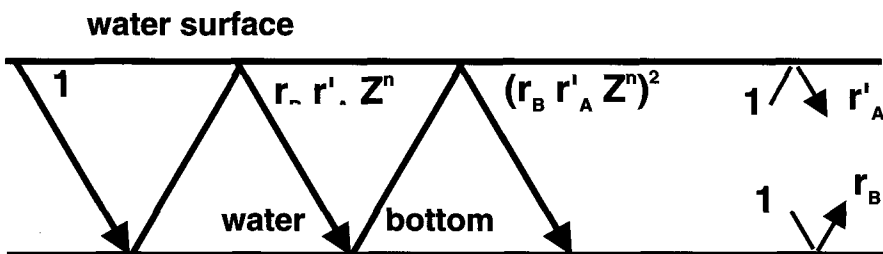


Fig. 4. Downgoing reverberation train.

Let  $r_B$  be the Fresnel reflection coefficient for a downgoing incident particle-velocity upgoing incident particle-velocity signal striking the water surface. Let the source produce a downgoing unit spike only. Let us suppose that we have a transducer that measures only down-traveling motion in the water layer. The signal received by the transducer represents the successive reverberations in the water layer. It is assumed that the two-way traveltime in the water layer is  $n$  time units. First let us consider the downgoing reverberation resulting from the downgoing unit spike source. The coefficient 1 occurs at time 0 and represents the initial downgoing source spike. The coefficient  $r_B r'_A$  occurs at discrete time  $n$  and represents the second downgoing spike, which has suffered a reflection at the bottom interface (reflection coefficient  $r_B$ ) and a reflection at the top interface (reflection coefficient  $r'_A$ ). The coefficient  $(r_B r'_A)^2$  occurs at discrete time  $2n$  and represents the third downgoing spike, which has suffered two reflections at the bottom and two reflections at the top. The coefficient  $(r_B r'_A)^3$  occurs at discrete time  $3n$  and represents the fourth downgoing spike, which has suffered three reflections at the bottom and three reflections at the top, etc. Thus the water-confined reverberation spike-train is of the form

$$M = 1 + r_B r'_A Z^n + (r_B r'_A Z^n)^2 + (r_B r'_A Z^n)^3 + \dots \quad (6)$$

This expression may be summed to give the following expression for the reverberation-producing filter

$$M = \frac{1}{1 - r_B r'_A Z^n} \quad (7)$$

The expression (7) holds for either downgoing or upgoing waves. The water-confined reverberation elimination filter would therefore have the form

$$H = 1 - r_B r'_A Z^n \quad (8)$$

This couplet is minimum-phase. Because the inverse of a minimum-phase wavelet is also minimum-phase, it follows that the water-confined reverberation spike-train (6) is minimum-phase.

## 7 Dual Fields

In 1742 Jean Le Rond d'Alembert (1717–1783) in the *Memoirs of the Berlin Academy* gave a solution to the wave equation that provides considerable insight into wave propagation phenomenon. He showed that a disturbance satisfying the wave equation is equal to the sum of two traveling waves, one of which travels downward and the other upward. The definition of the Fresnel reflection coefficient requires the consideration of both the particle-velocity attribute and the pressure attribute of the wave motion. Robinson (1983) discusses the various conventions used when dealing with pressure and particle velocity. The downgoing and upgoing waves of d'Alembert transport the energy

to and from a reflecting horizon, and the particle velocity and the pressure attributes of Fresnel determine the partition of energy at that horizon. Thus in seismic exploration there is an interplay of two dualities, namely the duality of the upgoing and downgoing waves of d'Alembert and the duality of the particle velocity and pressure of Fresnel. These two dualities represent a fundamental symmetry that should be respected in the acquisition of seismic data.

Let a dual sensor be placed at the top of layer  $\beta$ , that is, at a point just below interface  $\beta - 1$ . The sensor measures the particle-velocity signal and the pressure signal at that depth. Let  $D_\beta$  denote the downgoing component of the particle-velocity signal and let  $U_\beta$  denote the upgoing component of the particle-velocity signal. Similarly, let  $d_\beta$  denote the downgoing component of the pressure signal and let  $u_\beta$  denote the upgoing component of the pressure signal. Robinson (1984, p. 113) gives the d'Alembert equations

$$\begin{aligned} V_\beta &= D_\beta + U_\beta \\ p_\beta &= d_\beta + u_\beta \\ d_\beta &= \rho_\beta C_\beta D_\beta \\ u_\beta &= -\rho_\beta C_\beta U_\beta \end{aligned} \tag{9}$$

The third equation in (9) says that the downgoing pressure wave has the same polarity as the downgoing particle-velocity wave and that the two are related by a scale factor given by the acoustic impedance. The fourth equation in (9) says that the upgoing pressure wave has the opposite polarity as the upgoing particle-velocity wave and that the two are also related by the same scale factor. This convention is the same as the convention given by Berkhout (1987, pp. 199–200).

The d'Alembert equations (9) may be written in the form

$$D_\beta = \frac{V_\beta + (p_\beta / \rho_\beta C_\beta)}{2} \quad ; \quad U_\beta = \frac{V_\beta - (p_\beta / \rho_\beta C_\beta)}{2} \tag{10}$$

This form gives the expressions for the downgoing and upgoing particle-velocity waves. Given the geophone data alone, the separation of the wave motion into its component downgoing and upgoing waves can be made only under special circumstances. Given the hydrophone data alone, the separation of the wave motion into its component downgoing and upgoing waves can be made only under special circumstances. Given the geophone data and the hydrophone data, as well as the acoustic impedance at the location of the receiver, the d'Alembert equations (10) can be used to separate of the wave motion into its component downgoing and upgoing waves.

The d'Alembert equations are fundamental. For a dual sensor planted in a given layer, the geophone records the particle-velocity signal  $V_\beta$  and the hydrophone records the pressure signal  $P_\beta$ . If the acoustic impedance of that layer is known, the d'Alembert equations can be used to find the downgoing component  $D_\beta$  and the upgoing component



$U_\beta$  of the particle-velocity signal. In this paper we deal with these two components. Alternatively, one could deal with the corresponding d'Alembert equations for the two components  $d_\beta$  and  $u_\beta$  of the pressure signal.

The layered-earth system, as shown in Fig. 5, has  $n$  flat interfaces. A fictitious interface (interface 0) located on top is called the datum. Because the datum does not physically exist, its Fresnel reflection coefficient is zero, and its Fresnel transmission coefficient is one. The actual interfaces are numbered 1, 2, 3, ...,  $n$  in order of increasing depth.

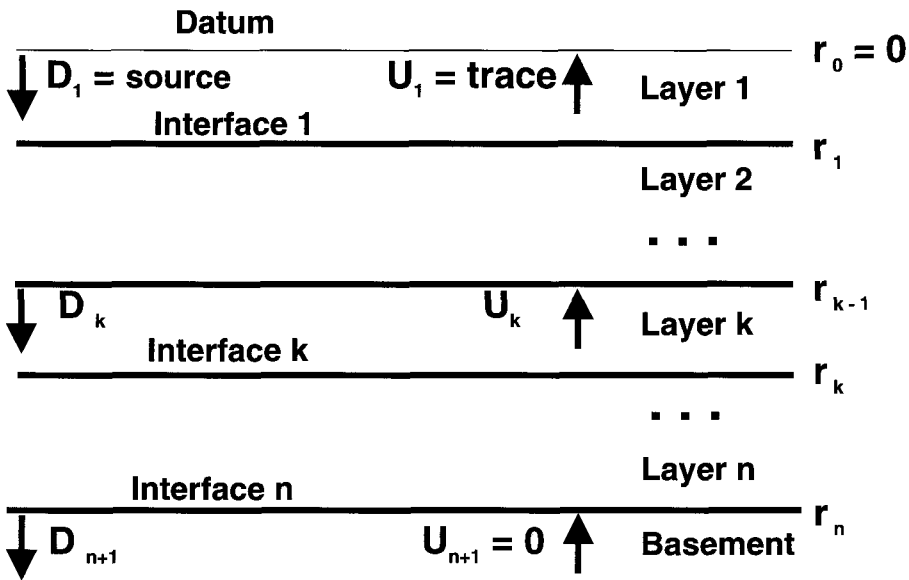


Fig. 5 The layered-earth model.

The respective reflection coefficients (for downgoing incident waves) are  $r_0, r_1, r_2, r_3, \dots, r_n$ . The individual interfaces separate the layers. The upper layer is called layer 1, the layer between interfaces 1 and 2 is called layer 2, the layer between interfaces 2 and 3 is called layer 3, and finally, the lower half-space is called layer  $n + 1$  or the basement. All waves are assumed to travel at normal incidence to the parallel interfaces. The waves are digitized with equal time spacing. For example, a time unit of 4 milliseconds might be used. The two-way travel time in each layer is taken to be this discrete unit of time. A wave in a given layer carries the number of the layer as a subscript. Although the waves exist throughout the layers, we will only be concerned with them as measured at points just below each interface. The convention is to measure a wave in layer  $k$  at the top of the layer; that is, at a depth point just below interface  $k - 1$ . A wave in layer 1 is measured at a depth point just below the datum.

## 8 System Reflection and Transmission Coefficients

Now let us turn our discussion of reflection and transmission coefficients from the case of just two layers to the case of a system of many layers. Instead of the so-called Fresnel reflection coefficient and Fresnel transmission coefficient we now obtain what are called the system reflection coefficient and the system transmission coefficient. The system coefficients are, of course, functions of the Fresnel coefficients that make up the system. The derivations given here are good for any system made up of multiple layers. More specifically, consider a stack of  $n$  layers in isolation from any other layers. For the purpose of exposition let the system have interfaces 1, 2, 3, ...,  $n$  with the respective Fresnel reflection coefficients  $r_1, r_2, r_3, \dots, r_n$ . The one-way downgoing system transmission factor and the one-way upgoing system transmission factor are defined respectively as

$$\sigma_n = \tau_n \cdots \tau_3 \tau_2 \tau_1 \quad ; \quad \sigma'_n = \tau'_n \cdots \tau'_3 \tau'_2 \tau'_1 \quad (11)$$

The fundamental polynomials  $P_n(Z)$  and  $Q_n(Z)$  are defined by the matrix equation (Robinson, 1967)

$$\begin{bmatrix} P_n^R & Q_n^R \\ Q_n & P_n \end{bmatrix} = \begin{bmatrix} Z & -r_n \\ -r_n & Z \end{bmatrix} \cdots \begin{bmatrix} Z & -r_2 \\ -r_2 & Z \end{bmatrix} \begin{bmatrix} Z & -r_1 \\ -r_1 & Z \end{bmatrix} \quad (12)$$

Here the reverse polynomials (with the superscript R for reverse) are given by

$$P_n^R(Z) = Z^n P_n(Z^{-1}) \quad ; \quad Q_n^R(Z) = Z^n Q_n(Z^{-1}) \quad (13)$$

The polynomial  $P_n(Z)$  is minimum-phase. The polynomials  $P_n(Z)$  and  $Q_n^R(Z)$  are each of degree  $n - 1$  whereas the polynomials  $Q_n(Z)$  and  $P_n^R(Z)$  are each of degree  $n$ . The relationship between the waves in layer  $n + 1$  to the waves in layer 1 are given by the matrix equation

$$\begin{bmatrix} D_{n+1} \\ U_{n+1} \end{bmatrix} = \frac{Z^{-n/2}}{\sigma'_n} \begin{bmatrix} P_n^R & Q_n^R \\ Q_n & P_n \end{bmatrix} \begin{bmatrix} D_1 \\ U_1 \end{bmatrix} \quad (14)$$

This equation is the matrix form of Lorentz transform (Robinson, 1967, 1982).

The reflection coefficient for the datum is zero. See Fig. 6 (left). Let the source be a downgoing unit spike at the datum. The system reflection coefficient  $R(Z)$  is the resulting upgoing escaping wave at the datum, and the system transmission coefficient  $T(Z)$  is the resulting downgoing escaping wave into the basement, that is,

$$R(Z) = U_1(Z) \quad ; \quad T(Z) = D_{n+1}(Z) \quad (15)$$

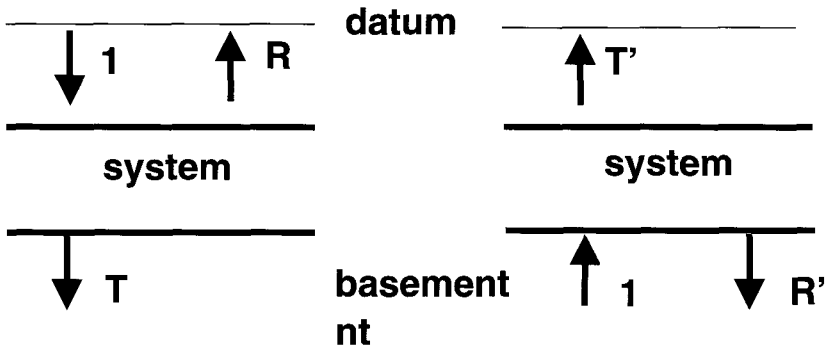


Fig. 6 (left). Downgoing source. (right) Upgoing source.

There is no upgoing wave in the lower half-space; that is,  $U_{n+1} = 0$ . Because the datum (interface 0) has a reflection coefficient of zero, the upgoing escaping wave is not reflected back into the system. Thus the downgoing wave at the datum is simply the initial unit source pulse; that is,  $D1 = 1$ . Thus equation (14) becomes

$$\begin{bmatrix} T \\ 0 \end{bmatrix} = \frac{Z^{-n/2}}{\sigma'_n} \begin{bmatrix} P_n^R & Q_n^R \\ Q_n & P_n \end{bmatrix} \begin{bmatrix} 1 \\ R \end{bmatrix} \tag{16}$$

The solution of this equation gives the system reflection coefficient and the system transmission coefficient as

$$R(Z) = \frac{-Q_n(Z)}{P_n(Z)} \quad ; \quad T(Z) = \frac{\sigma_n Z^{n/2}}{P_n(Z)} \tag{17}$$

Now let the source be an upgoing unit spike just below interface  $n$ . See Fig. 6 (right). The system reflection coefficient  $R'(Z)$  is the resulting downgoing wave just below interface  $n$ , and the system transmission coefficient  $T'(Z)$  is the resulting upgoing wave at the datum; that is,

$$R'(Z) = D_{n+1}(Z) \quad ; \quad T'(Z) = U_1(Z) \tag{18}$$

Thus

$$\begin{bmatrix} R' \\ 1 \end{bmatrix} = \frac{Z^{-n/2}}{\sigma'_n} \begin{bmatrix} P_n^R & Q_n^R \\ Q_n & P_n \end{bmatrix} \begin{bmatrix} 0 \\ T' \end{bmatrix} \tag{19}$$

The solution of this equation gives the upgoing system coefficients

$$R'(Z) = \frac{Q_n^R(Z)}{P_n(Z)} \quad ; \quad T'(Z) = \frac{\sigma_n' Z^{n/2}}{P_n(Z)} \quad (20)$$

The system transmission coefficients  $T$  and  $T'$  are each minimum-phase.

## 9 Reverberations Between Two Systems

Choose a system with  $n$  interfaces, numbered

$$1, 2, 3, \dots, \alpha - 1, \alpha, \alpha + 1, \dots, n \quad (21)$$

Interface 1 is the surface of the ground or the surface of the water as the case may be. The source is in layer  $\alpha$ , which is the layer between interface  $\alpha - 1$  and interface  $\alpha$ . More specifically, the source is at the top of layer  $\alpha$ , so the source is just below interface  $\alpha - 1$ . Layer  $\alpha$  is called the source layer. The lowermost interface is interface  $n$ . All the material below interface  $n$  is referred to as basement rock. See Fig. 7.

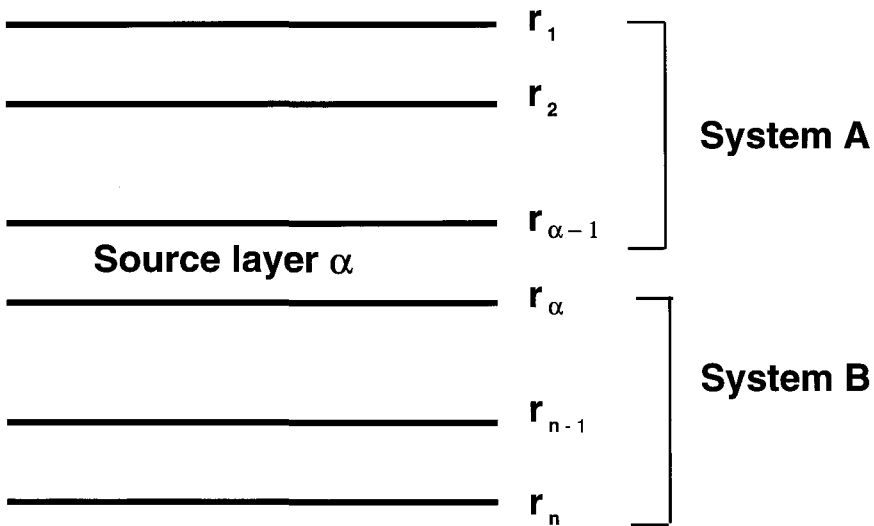


Fig. 7. The two component systems separated by the source layer.

The Fresnel reflection coefficients of the system are given by  $r_1, r_2, \dots, r_n$ . Break the given system into two component systems, denoted by A and B. System A contains all the interfaces above the source. System B contains all the interfaces below the source and above the basement rock. System A has the series of reflection coefficients  $r_1, r_2, \dots, r_{\alpha-1}$ . System B has the series of reflection coefficients  $r_\alpha, r_{\alpha-1}, \dots, r_n$ .

The fundamental polynomials can be found from the reflection coefficients for each system in isolation from the other systems. The fundamental polynomials for any system

are identified by the subscript. For example,  $P_A$  and  $Q_A$  are the fundamental polynomials for the upper system A. Similarly  $P_B$  and  $Q_B$  are the fundamental polynomials for the lower system B. From the fundamental polynomials, the system reflection and transmission coefficients can be found. For example, we have

$$R'_A(Z) = \frac{Q_A^R(Z)}{P_A(Z)} \quad ; \quad R_B(Z) = -\frac{Q_B(Z)}{P_B(Z)} \tag{22}$$

System A (namely, the system above the source) has the system reflection coefficient  $R'_A$  for waves striking the system from below. System B (namely, the system below the source) has the system reflection coefficient  $R_B$  for waves striking the system from above. The source is excited, producing a downgoing particle-velocity signal  $S$  as well as an upgoing particle-velocity signal  $-S$ . The ghost-producing filter can be obtained from equation (4) by replacing  $r'_A Z^n$  by  $R'_A$ . Thus the ghost-producing filter for this case is

$$G = 1 - R'_A \tag{23}$$

The source produces reverberations in the source layer, that is, in layer  $\alpha$ . The mathematical structure is the same as given for the case of reverberations between two interfaces. However, now system coefficients must be used instead of Fresnel coefficients. Thus the system reflection coefficient  $R'_A$  is used for the layers above the source. Likewise, the system reflection coefficient  $R_B$  is used for the layers below the source. The reverberation-producing filter can be obtained from equation (7) by replacing  $r_B r'_A Z^n$  by  $R_B R'_A$ . Thus the reverberation-producing filter for this case is

$$M = \frac{1}{1 - R_B R'_A} \tag{24}$$

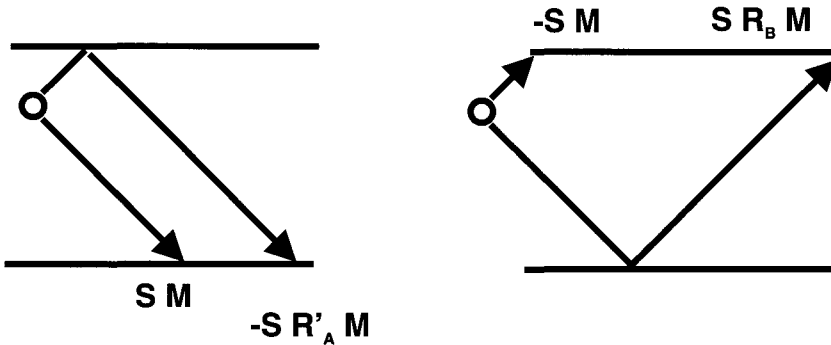


Fig. 8. (left) Components of  $D_\alpha$ . (Right) Components of  $U_\alpha$ .

It follows that the downgoing particle-velocity wave in the source layer is as shown in Fig. 8 (left), which is

$$D_\alpha = S G M = \frac{S (1 - R'_A)}{(1 - R_B R'_A)} = S M - S R'_A M \quad (25)$$

The upgoing reflected wave  $U_\alpha^r$  in the source layer is the reflection of the downgoing particle-velocity wave from System B. This upgoing reflected wave is given by

$$U_\alpha^r = R_B D_\alpha = S R_B G M \quad (26)$$

The direct upgoing wave  $U_\alpha^d$  in the source layer due to the source is

$$U_\alpha^d = -S \quad (27)$$

The entire upgoing wave is sum of the direct upgoing wave (27) plus the upgoing reflected wave (26) is as shown in Fig. 8 (right), which is

$$U_\alpha = U_\alpha^d + U_\alpha^r = -S + S R_B G M = -S M + S R_B M \quad (28)$$

## 10 Dual-Field Deconvolution

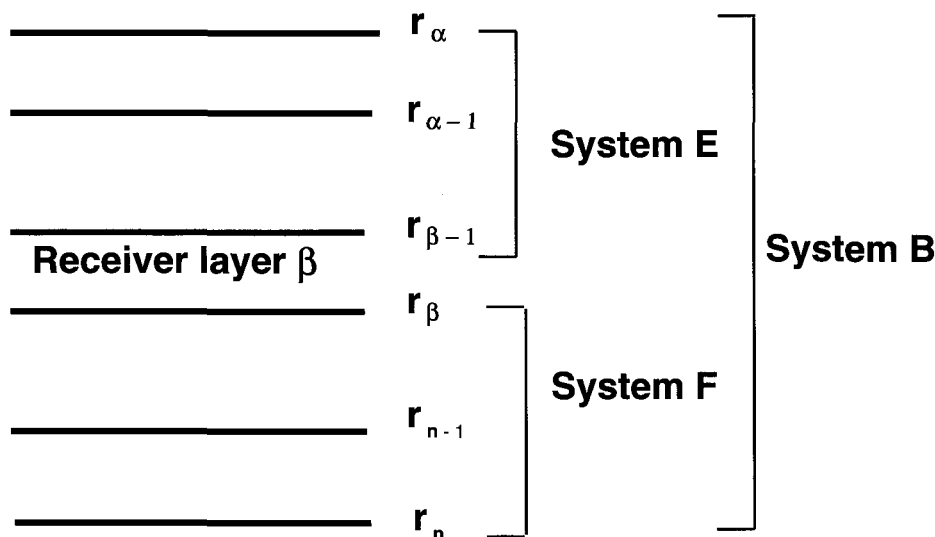


Fig. 9. The two component systems making up system B with the separation at the source layer.

Fig. 9 shows the receiver in layer  $\beta$ . The receiver (a dual sensor) could be above the source, or at the source, or below the source. However, we treat only the case where the receiver is strictly below the source, so  $\beta > \alpha$ . Recall that system B contains all the

interfaces below the source; that is, system B has the series of reflection coefficients  $r_\alpha, r_{\alpha+1}, \dots, r_n$ . Break system B into two systems, where the top system E is made up of the interfaces between source and receiver, and the lower system F is made up of the interfaces between the receiver and the basement rock. Thus system E has Fresnel reflection coefficients  $r_\alpha, r_{\alpha+1}, \dots, r_{\beta-1}$  and system F has Fresnel reflection coefficients  $r_\beta, r_{\beta+1}, r_n$ .

Let the downgoing and upgoing waves in the receiver layer be  $D_\beta$  and  $U_\beta$  respectively. They are related by

$$U_\beta = R_F D_\beta \quad (29)$$

where  $R_F$  is the reflection impulse response of the layers below the receiver. Consider subsystem E alone. From the Lorentz transform (14) we have

$$\begin{bmatrix} D_\beta \\ U_\beta \end{bmatrix} = \frac{Z^{-(\beta-\alpha)/2}}{\sigma'_E} \begin{bmatrix} P_E^R & Q_E^R \\ Q_E & P_E \end{bmatrix} \begin{bmatrix} D_\alpha \\ U'_\alpha \end{bmatrix} \quad (30)$$

From equations (25), (26) and (30), we obtain the following expressions for the waves  $U_\beta$  and  $D_\beta$  in the receiver layer

$$\begin{aligned} D_\beta &= \frac{Z^{-(\beta-\alpha)/2}}{\sigma'_E} (P_E^R + Q_E^R R_B) S G M \\ U_\beta &= \frac{Z^{-(\beta-\alpha)/2}}{\sigma'_E} (Q_E + P_E R_B) S G M \end{aligned} \quad (31)$$

Equation (29) may be written as

$$R_F = \frac{U_\beta}{D_\beta} \quad (32)$$

System F, namely, the system between the receiver and the basement rock, contains the reflection coefficients of interest in exploration. The near-surface reflection coefficients, that is, those of the interfaces above the receiver, give rise to the reverberations and ghosts that we want to eliminate. The dual sensor measures both the particle velocity signal  $V_\beta$  and the pressure signal  $p_\beta$  in layer  $\beta$ . In addition the acoustic impedance  $p_\beta C_\beta$  at the receiver layer must be obtained or estimated.

The Einstein deconvolution method can be described in two steps. The first step is to convert the particle velocity and pressure signals into the downgoing wave  $D_\beta$  and the upgoing wave  $U_\beta$  in layer  $\beta$ . In order to carry out this step the d'Alembert equations (10) are used. The second step is to deconvolve the upgoing wave by the downgoing wave. In

order to carry out this step the right hand side of equation (32) is used. The result of Einstein deconvolution is the unit-impulse reflection response  $R_F$  of system F.

Another expression for  $R_F$  can be obtained as follows. From equations (31) we have

$$\frac{U_\beta}{D_\beta} = \frac{(Q_E / P_E) + R_B}{(P_E^R / P_E) + (Q_E^R / P_E) R_B} \quad (33)$$

The quantity  $\Omega_E = P_E^R / P_E$  is an all-pass system. The all-pass system adjusts the phase spectrum without changing the amplitude spectrum. Because

$$R_F = U_\beta / D_\beta, \quad R_E = -Q_E / P_E, \quad R'_E = Q_E^R / P_E \quad (34)$$

equation (33) becomes

$$R_F = \frac{-R_E + R_B}{\Omega_E + R'_E R_B} \quad (35)$$

The novelty of expression (35) rests in the appearance of the all-pass system  $\Omega_E$ . In summary, the deconvolution of the upgoing wave recorded at the receiver by the downgoing wave recorded at the receiver gives the unit-impulse reflection response  $R_F$  of the subsystem below the receiver. The Einstein deconvolution process strips away the multiples and ghosts caused by upper system. It should be emphasized that the Einstein deconvolution process also strips away the unknown signature wavelet  $S$ .

## 11 Dynamic Deconvolution

Dynamic deconvolution (Robinson, 1967, 1975) is a form of layer stripping. See also Darby and Neidell (1967). Let us consider the simplest case of layer stripping, namely, the case of layer stripping for a downgoing unit-spike source incident on the surface. The resulting upgoing wave measured on the surface represents the reflection unit-impulse response  $R$  of the medium. Because the wave motion incident on the surface is a downgoing unit spike, it follows that the first primary reflection is equal to the reflection coefficient of the surface. Knowing the first reflection coefficient we can strip off the layer between the surface and the next interface. The next interface now appears as the new surface, and thus the layer-stripping process can be repeated to obtain the next reflection coefficient. This layer-stripping process can be repeated over and over to determine all the reflection coefficients, one by one, as the scheme probes deeper and deeper into the earth. Immediately one sees the weakness of this layer stripping scheme as well as other similar approaches. All layer-stripping schemes assume that the unit-impulse reflection response is available, that is, the response of the layered earth to a unit spike. Certainly, in seismic acquisition, a perfect unit spike source is never available. In actuality, the source produces some kind of signature  $S$ , and an accurate estimate or measurement of this signature is difficult to obtain at best. The making of such measurements is fraught with troubles. Hagedoorn (1964) discusses the difficulties of



detecting the first arrival against a background of seismic disturbances due to sources other than the explosion. Because dynamic deconvolution is a nonlinear process, an accurate estimate is mandatory. Because of this shortcoming, dynamic deconvolution and other similar layer stripping methods have found little or no use in practice. Clearly, for dynamic deconvolution to work, a superior estimate of the unit-impulse reflection response is needed.

As we have seen Einstein deconvolution removes the (unknown and not necessarily minimum-phase) source signature as well as the reverberations and ghosts due to the layers above the receiver. The resulting deconvolved record is the unit-impulse reflection response of the geological interfaces below the receiver. Thus the output of Einstein deconvolution is precisely the input required for dynamic deconvolution. The output of dynamic deconvolution is the reflection coefficient series. Thus Einstein deconvolution followed by dynamic deconvolution yields a good estimate of the series of reflection coefficients of the interfaces below the receiver.

## 12 Conclusions

A receiver made up of dual geophone-hydrophone sensors measures two attributes of the wavefield. One attribute is particle velocity and the other attribute is pressure. In this paper, the dual-sensor receiver is buried at a level below the level of the buried source. Einstein deconvolution, which requires the dual-sensor data, removes all the reverberations and ghosts due to interfaces above the receiver. Einstein deconvolution also removes the unknown source signature in the same operation. The resulting deconvolved seismogram is the unit-impulse reflection response that would be produced as if there were no layers at all above the buried receiver. If desired, dynamic deconvolution can then be performed on the unit-impulse reflection response obtained by Einstein deconvolution. The output of the dynamic deconvolution process is the sequence of reflection coefficients for the interfaces below the receiver. Because its theory is based on the Z-transform, it follows that Einstein deconvolution operates under the same limitations as predictive deconvolution, namely flat horizontally layers subjected to seismic compressional waves at normal incidence. As a result, the limitations for Einstein deconvolution can be addressed by the same means as is done for predictive deconvolution.

The common goal of both predictive deconvolution and Einstein deconvolution is to obtain the reflection coefficient series as the deconvolved signal. Predictive deconvolution requires the unknown source signature to be minimum-phase, whereas Einstein deconvolution does not. Both predictive (spiking) deconvolution and Einstein deconvolution carry out the deconvolution process on the upgoing signal. Both predictive deconvolution and Einstein deconvolution have the same deconvolution operator, namely the inverse of the downgoing signal. Thus these two methods of deconvolution look like each other. The difference is in the fundamental assumptions that determine the way the deconvolution operator is obtained. The small white reflectivity hypothesis allows the

predictive deconvolution operator to be computed by least-squares from the upgoing signal. In addition the small white reflectivity hypothesis eliminates the necessity for the final dynamic-deconvolution step. Predictive deconvolution has the advantage of nearly fifty year's usage. It is robust and stable in the presence of noise. Einstein deconvolution has the advantage the small white reflectivity hypothesis is not required. In this way Einstein deconvolution is more general. However Einstein deconvolution is more sensitive to noise. Ideally both methods can be used in conjunction with one another to obtain better results.

### 13 Acknowledgment

Grateful acknowledgement is given to Prof. C. H. Chen of the University of Massachusetts Dartmouth for his support of this work, and for the many valuable scientific ideas he has given to the author over the past thirty years.

### References

1. Aminzadeh F. 1984. "Derivation of layer parameters of an elastic medium from reflection coefficient matrices." *Geophysical Prospecting*, 32, 819–827.
2. Adair R. D., Orcutt J. A., and Farrell W. E. 1988. "Infrasonic Seismic and Acoustic measurements in the deep ocean." *IEEE Journal of Ocean Engineering*, 13, 245–253.
3. Barr F. J. and Sanders J. I. 1989. "Attenuation of Water-Column Reverberations Using Pressure and Velocity Detectors in a Water-Bottom Cable." *SEG Expanded Abstracts*, CD ROM.
4. Barr F. J. 1990. System for attenuating water-column reverberations. U. S. Patent Office, No. 4979150.
5. Barr F. J., Paffenholz J., and Rabson W. 1966. "The dual-sensor ocean-bottom cable method." *SEG Expanded Abstracts*, CD ROM.
6. Barr F. J. 1997. "Dual sensor OBC technology." *The Leading Edge*, 16, 45–47.
7. Berkhout A. J. 1987. *Applied Seismic Wave Theory*. Elsevier, Amsterdam.
8. Berni A. J. 1984. *Marine seismic system*. U. S. Patent Office, No. 4437175.
9. Berni A. J. 1985. *Marine seismic system*. U. S. Patent Office, No. 4520467.
10. Canales L. L. and Bell M. L. 1996, "Ghost attenuation using dual sensor cable." *SEG Expanded Abstracts*, 1591–1594.
11. Cowles C. S. 1979. *Combination geophone-hydrophone*. U. S. Patent Office, No. 4134097.
12. Darby E. K. and Neidell N. S. 1967. "Application of dynamic programming to the problem of plane wave propagation in a layered medium." *Geophysics*, 31, 1037–1048.
13. Dragoset B. and Barr F. J. 1994. "Ocean-bottom cable dual-sensor scaling." *SEG Expanded Abstracts*, 857–860.

14. Einstein A. 1905. "Zur Elektrodynamik bewegter Körper." *Annalen der Physik*, 17, 891–921.
15. Gal'perin E. I. 1971. *Vertical Seismic Profiling*. Society of Exploration Geophysicists, Tulsa.
16. Gutowski P. R. and Treitel S. 1987. "The generalized one-dimensional synthetic seismogram." *Geophysics*, 52, 589–605.
17. Hagedoorn J. G. 1964. "The elusive first arrival." *Geophysics*, 29, 806–813.
18. Haggerty P. E. 1956. *Method and apparatus for canceling reverberations in water layers*. U. S. Patent Office, No. 2757356.
19. Loewenthal D. 1975. "Theoretical uniqueness of the magnetotelluric inverse problem for equal penetration discretized models." *Geophysical Journal Roy. Ast. Soc.* 43, 897–903.
20. Loewenthal D. and Jakubowicz H. 1983. "Wave equation-based source signature deconvolution." *SEG Expanded Abstracts*, 430–431.
21. Loewenthal D., Lee S. S. and Gardner G. H. F. 1985. "Deterministic estimation of a wavelet using impedance type technique." *Geophysical Prospecting*, 33, 956–969.
22. Loewenthal D. and Shtivelman V. 1989. "Source signature estimation using fictitious source and reflector." *Geophysics*, 54.
23. Loewenthal, D., 1991. "Proper boundary conditions in reflection seismograms." *J. Geophys Res.* 96, 20321–20324.
24. Loewenthal D., and Stoffa P. L. 1991. "Synthetic seismograms by dereverberating sources." *J. Acoust. Soc. Am.* 90, 1101–1105.
25. Loewenthal D., 1994. "On dual field measurements using geohydrophones." *SEG Expanded Abstracts*, 861–864.
26. Paffenholz J. and Barr J. 1995. "An improved method for deriving water-bottom reflectivities for processing dual-sensor ocean-bottom cable data." *SEG Expanded Abstracts*, CD ROM.
27. Pavey G. M. 1966. *Method and underwater streamer apparatus for improving the fidelity of recorded seismic signals*. U. S. Patent Office, No. 3290645.
28. Pavey G. M. 1967. *Underwater detection streamer apparatus for improving the fidelity of recorded seismic signals*. U. S. Patent Office, No. 3299397.
29. Robinson E. A. 1966. "Multichannel Z-transforms and minimum-delay." *Geophysics* 31, 482–500.
30. Robinson E. A. 1967. *Multichannel Time Series Analysis with Digital Computer Programs*. Emerson-Adams Press, Inc., Boca Raton, Florida.
31. Robinson E. A. 1975. "Dynamic predictive decomposition." *Geophysical Prospecting* 23, 779–797.
32. Robinson E. A. 1982. "Spectral approach to geophysical inversion by Lorentz, Fourier, and Radon Transforms." *Proceedings of the IEEE*, 70, 1039–1054.
33. Robinson E. A. 1983. *Seismic Velocity Analysis and the Convolutional Model*. Prentice-Hall, Englewood Cliffs, NJ, and D. Reidel Publishing Co., Dordrecht.

34. Robinson E. A. 1984. *Seismic Inversion and Deconvolution. Part A. Classical Methods*. Elsevier, Amsterdam.
35. Robinson E. A. 1985. "Seismic time-invariant convolutional model." *Geophysics*, 50, 2742–2751.
36. Robinson, E. A. 1999. *Seismic Inversion and Deconvolution: Part B. Dual Sensor Technology*, Pergamon (An imprint of Elsevier), Amsterdam, 348 pages.
37. Robinson E. A. and Treitel S. 1996. "Net downgoing energy and the minimum-delay property of downgoing waves." *Seismic Source Signature Estimation and Measurement*, Society of Exploration Geophysicists, Tulsa.
38. Ruehle W. H. 1984. *Pressure and velocity detectors for seismic exploration*. U. S. Patent Office, No. 4486865.
39. Shtivelman V. and Loewenthal D. 1988. "Source wavelet estimation by upward extrapolation." *Geophysics* 53, 158–166.
40. Shtivelman V. and Loewenthal D. 1989. "Construction of the generalized one-dimensional synthetic seismograms by a three-step extrapolation procedure." *Geophysics*, 54, 1050–1053
41. Schneider W. A. and Backus M. M. 1964. "Ocean-bottom seismic measurements of the California coast." *Journal of Geophysical Research*, 69, 1135–1143
42. Sun C. 1997. "Separation of vertical cable up-going and down-going wave fields." *SEG Expanded Abstracts*, CD ROM.
43. Szaraniec E. 1985. "On direct recovery of the impulse response." *Geophysical Prospecting*, 33, 498–502–969.
44. White J. E. 1965. *Seismic Waves: Radiation, Transmission, and Attenuation*. McGraw-Hill, NY.
45. Ziolkowski A. M., Lerwill W. E., March D. W., and Peardon L. G. 1980. "Wavelet deconvolution using a source scaling law." *Geophysical Prospecting*, 28, 872–901.

This page is intentionally left blank

## CHAPTER 20

### EDGE-PRESERVING SMOOTHING FOR ENHANCING 3-D SEISMIC IMAGES

Yi Luo, Maher Marhoon, Saleh Al-Dossary and Mohammed Alfaraj

*Saudi Aramo, Dhahran, 31311, Saudi Arabia*  
*E-mail: yi.luo@aramco.com*

We present a new noise-reduction method, named Edge-Preserving Smoothing (EPS), to reduce noise in 3-D seismic data. The EPS method simply attempts to suppress random noise and some acquisition artifacts in seismic data while preserving sharp boundaries or edges; these edges often correspond to important geological features, such as faults, fractures and channels. By applying EPS as a pre-processing step before running algorithms to detect edges in seismic data, we have obtained seismic edge-detection results with much-improved S/N ratio and resolution.

#### 1 Introduction

Suppressing random noise is an important pre-processing step before applying seismic edge-detection (or coherence-cube) algorithms (Bahorich and Farmer, 1995; Gersztenkorn and Marfurt, 1996; Luo et al., 1996; and Luo et al., 2002). The reason is that reflection data near faults/fractures are usually more complex and noisier than in other areas, and most edge-detection algorithms, which attempt to highlight local rapid changes in seismic data, are sensitive to noise.

Usually, Prediction-Error Filtering (PEF) or  $f_x$ -deconvolution (Claerbout, 1998) is used to precondition the data before attempting edge detection. Although PEF has been very successful in many areas, it tends to inadequately remove noise in areas close to faults and fractures, where reflection signals are usually not highly predictable.

A simple and robust alternative method for reducing noise is to use moving-window or running-average smoothing filters. Unlike the PEF method, such smoothing does not strongly depend on the predictability of signals. The drawback of this smoothing, however, is that it tends to blur sharp edges associated with faults and channels, which one would ideally want to enhance with seismic-edge detection algorithms.

The EPS method attempts to resolve the conflict between noise reduction and edge degradation. This method is obtained by a simple modification of the running-average smoothing method. Such modification enables EPS to suppress noise while keeping sharp edges intact.

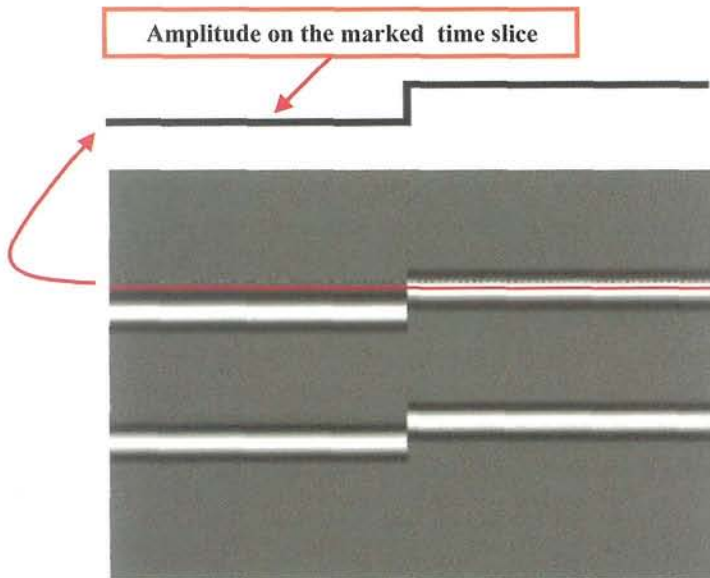
The EPS method was proposed (Nagao, 1980) to reduce noise in 2-D images in the remote sensing field. We extended this method to 3-D images, which represent subsurface structures of the earth, and applied extended EPS method to seismic data.

## 2 EPS Method

In principle, EPS looks for the most homogeneous fragment around each sample in an input data set and assigns the average value of the selected fragment to that sample. The method was proposed for reducing noise in 2-D photographs. In this section, we will extend the method into 1-D function for simplifying the description of the basic idea, and 3-D volume of subsurface seismic images.

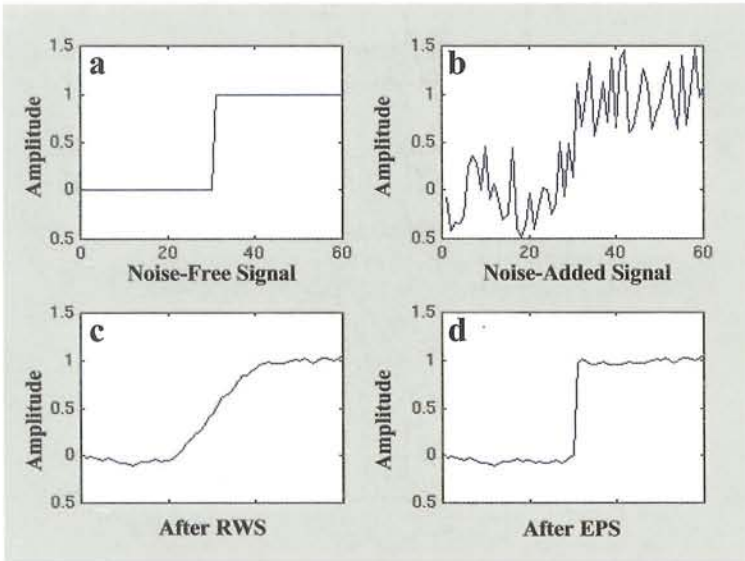
### 2.1 1-D EPS Method

Figure 1 shows a vertical synthetic seismic section with an ideal fault in it. The amplitudes on a time slice, which cross the faults, form a step function. This step function is used in Figure 2 to illustrate the concept and benefits of EPS. Figure 2A displays a noise-free step function, and 1B shows the same function after adding random noise.



**Figure 1: Ideal fault. The amplitudes on a time slice form a step function.**

Applying a 21-point running-average smoothing filter to the noisy function of Figure 2B yields the result shown in Figure 2C. The random noise is obviously reduced in Figure 2C, but the sharp step is severely altered. Figure 2D depicts the result after applying a 21-point EPS operator to the noisy step function in Figure 2B. It is clear from Figure 2D that the sharp edge/step is well preserved while the noise is reduced successfully.



**Figure 2: Concept of EPS. Input step function (A); noise-added step function (B); result after regular smoothing (C) and result after EPS(D).**

To simply clarify how the EPS method works, let us take a 3-point EPS operator instead of 21 points as an example. In this case, for any given output location at index  $i$ , we first calculate standard deviations (Caswell, 1995) for the following five shifted windows:

window 1: (  $A_{i-2}$ ,  $A_{i-1}$ ,  $A_{i+0}$  ),

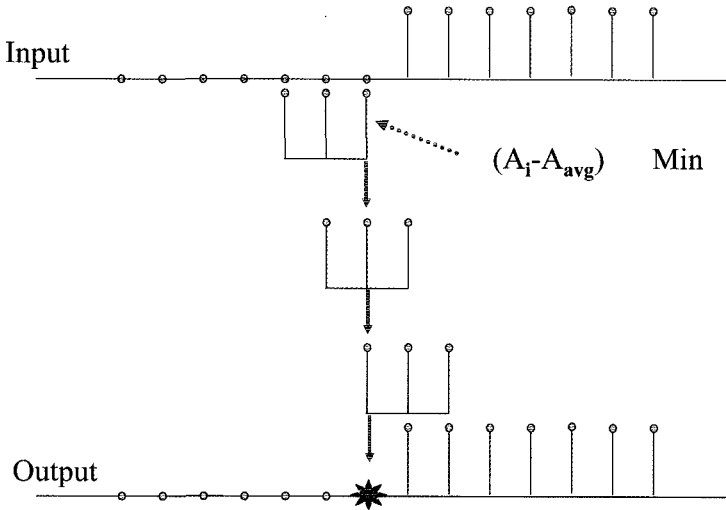
window 2: (  $A_{i-1}$ ,  $A_{i+0}$ ,  $A_{i+1}$  ),

window 3: (  $A_{i+0}$ ,  $A_{i+1}$ ,  $A_{i+2}$  ),

Here,  $A_i$  represents the amplitude of the  $i$ th sample in the input data. Next, we select the window which has the minimum standard deviation, calculate the average over the selected window and assign the average as output at the  $i$ th output location. Repeating this process for all the output locations will yield the result shown in Figure 2D (assuming a 21-point window is used). For a  $N$ -point window, we will get  $N$  elective average values for each output location, the one calculated on the window with minimum standard deviation will be chosen and output.

This procedure described above is illustrated in Figure 3. In the figure, the current output location under working is indicated by the "star" in the output signals. Three elective windows are shown in between the input and output signals. The average value of the top window will be used as the output due to the top window has minimum zero standard deviation.

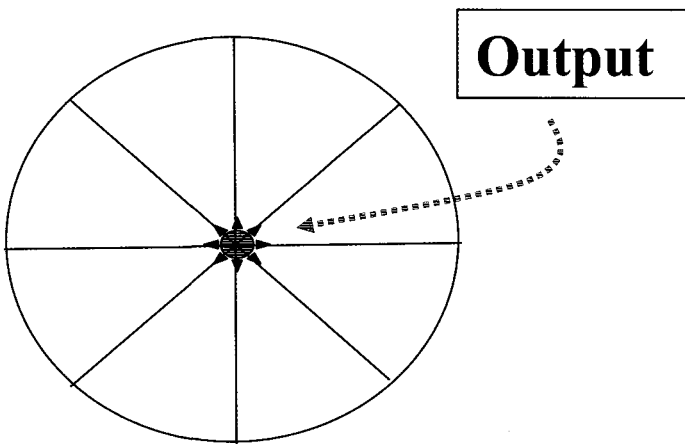




**Figure 3: Discrete operators of 1-D EPS.**

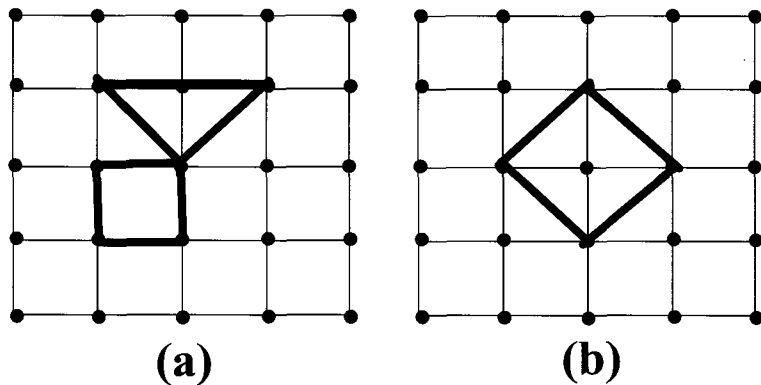
## 2.2 2-D EPS Method

This general scheme can be readily extended to 2-D and 3-D cases. For any output location  $(x,y)$  in the 2-D case, for example, we divide its vicinity into small fragments, and calculate the standard deviation of the input data for each fragment, respectively (Figure 4). The fragment corresponding to the smallest standard deviation will be selected, and its average value is used as the output for the location  $(x,y)$ .

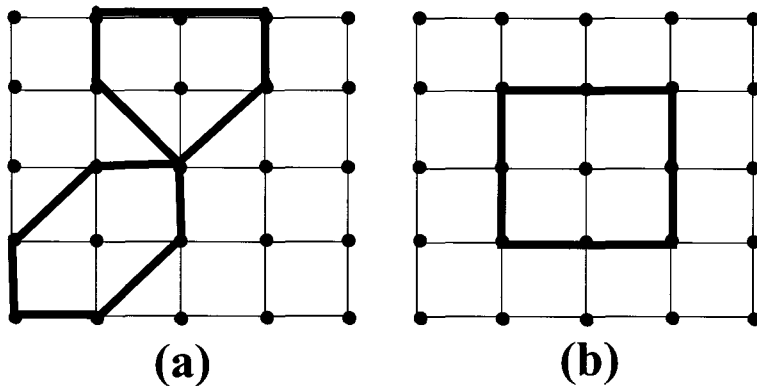


**Figure 4: Concept of 2-D EPS.**

Various numerical approaches can be adapted for implementing the concept shown in Figure 4. Figures 5 and 6 show a discrete implementation of 3X3 window a 5X5 window respectively. Larger windows (7X7, 9X9 etc.) can be used for noisier input data, and the vicinity of a output point can be divided into more then nine pieces as we did in Figures 5 and 6.

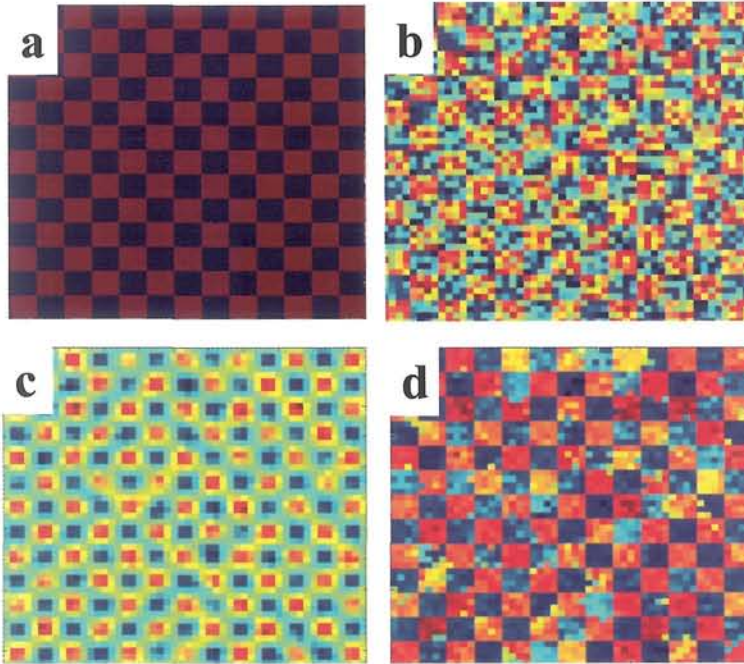


**Figure 5:** Discrete realization of 2-D EPS operator in a 3X3 window. (a) One triangle and one square operator. There should be four triangle and four square operators in total (not depicted). (b) Central operator.



**Figure 6:** Discrete realization of 2-D EPS operator in a 5X5 window. (a) One pentagonal and one hexagonal operator. There should be four pentagonal and four hexagonal operators in total (not depicted). (b) Central operator.

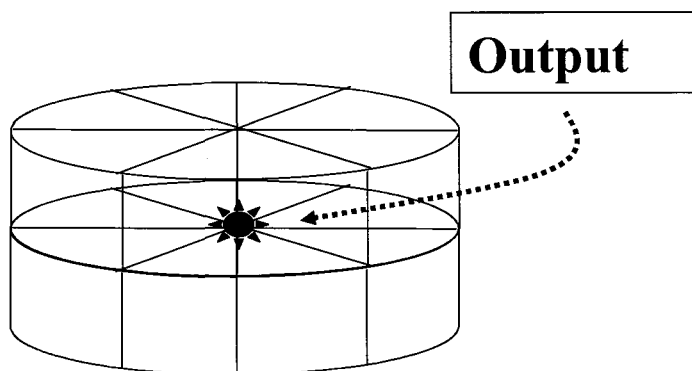
Figure 7 illustrates how the EPS can remove random noise without hurting sharp edges in the data. Figures 7A and 7B show a 2-D input array without and with noise, respectively. Figure 7C displays the output array after applying a 5X5 running window to the noise-added array shown in Figure 7B, while Figure 7D depicts the result of the 5X5 2-D EPS. It is evident that the EPS can preserve sharp boundaries while suppressing noise



**Figure 7:** (a) 2-D input array without noise. (b) 2-D input array after adding random noise. (c) 2-D output array after applying 2-D running-window smooth. Blurring between adjacent box functions is obvious. (d) 2-D output array after applying 2-D EPS smooth. The boundaries between adjacent box functions are well preserved while the noise is well suppressed.

### 2.3 3-D EPS Method

Seismic data used in interpretation nowadays are often 3D volume, which manifest the structures of subsurface in 3D sense. Applying the method discussed above to 3-D seismic data, we have to extend EPS from 2D to 3D. Figure 8 illustrates the concept of 3-D EPS.



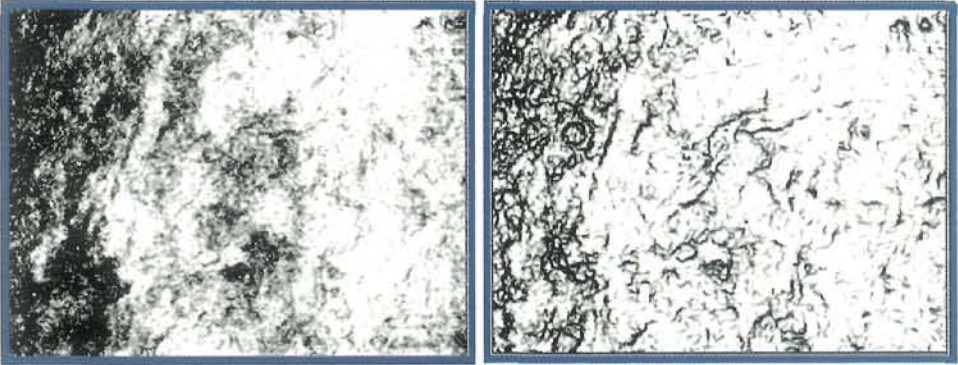
**Figure 8: Concept of 3-D EPS.**

In Figure 8, the surrounding area of a output point is divided into number of wedge-shape pieces (imaging cutting a thick pizza). For every wedge, a average and a deviation value can be calculated. The average value generated in the wedge with minimum standard deviation will be assigned as the output at the point under working. The discrete implementation for 3-D EPS can be the same as for 2-D (Figures 5-6), except each sample in 2-D EPS is a scalar (a value), but a vector in 3-D. The vector consists of a few samples (say 11 samples ) in the vertical direction; or it is a segment of a seismic trace.

### 3 Field Data Demonstrations

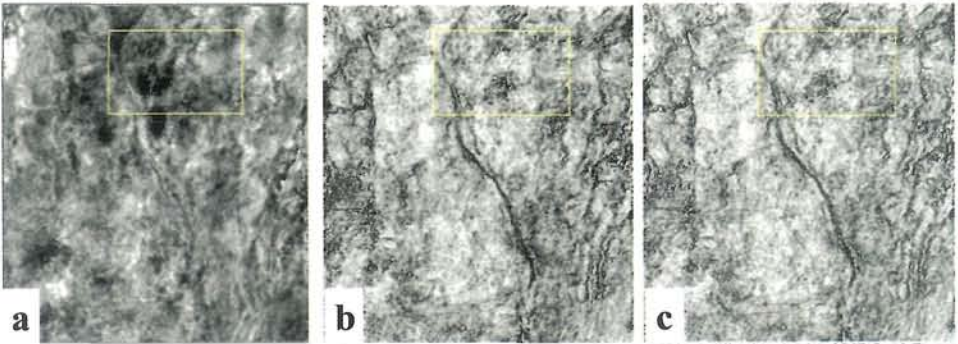
We test the EPS method by applying an edge-detection algorithm (Luo et al., 2002) to seismic data with and without EPS applied. In these experiments, we held constant all the input parameters used in the edge-detection algorithm, so that any differences in the results would be solely attributed to the EPS algorithm. Besides us, Kurt Marfurt et. al (2002) had applied this method to real cases, and some good examples can be found in his work.

Figure 9 shows time slices produced by applying the edge-detection algorithm to a 3-D post-stack seismic cube. Figure 9A is the edge-detection result without EPS being applied to the 3-D input cube, while Figure 9B is obtained with EPS being applied. Since noise in the input data is significantly reduced by the EPS, many lineation features stand out which are hardly seen in Figure 9A. However, if these features are true or not remains to be a question.



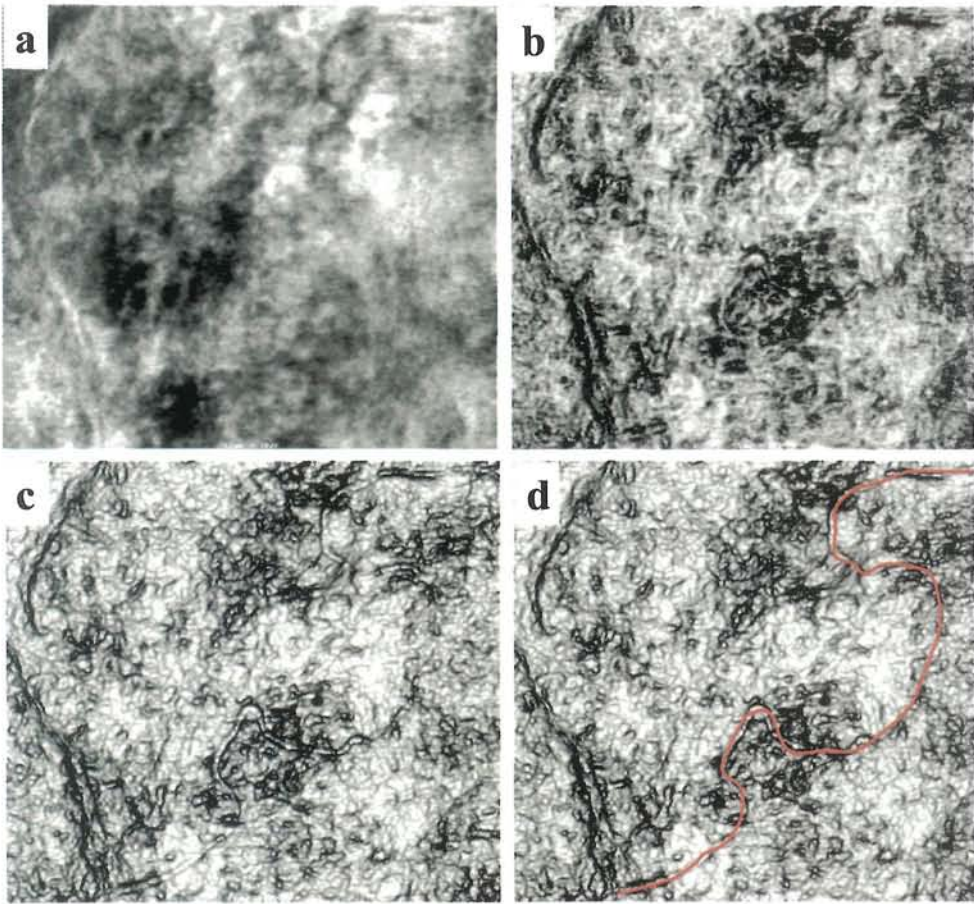
**Figure 9:** Time slice produced by applying edge detection to input data without (left) and with (right) EPS applied.

To answer this question, Figures 10 and 11 are used within which, as you will see, their imbedded channels and faults can be recognized with certain. Figure 10A shows an amplitude time slice extracted from a 3-D seismic cube,. Figure 10B displays the time slice of the edge-detection result without applying EPS, while 5C shows the edge-detection result with EPS at a pre-processing stage. Figure 11 is a close-up of Figure 10. In Figure 11D, we marked two channels and two fault zones, these features are clearer and sharper in Figure 11C than in Figure 11B.



**Figure 10:** (a) Amplitude time slice of input data. (b) Time slice produced by applying edge-detection algorithm to input data without EPS. (c) Time slice produced by applying edge-detection algorithm to input data with EPS.



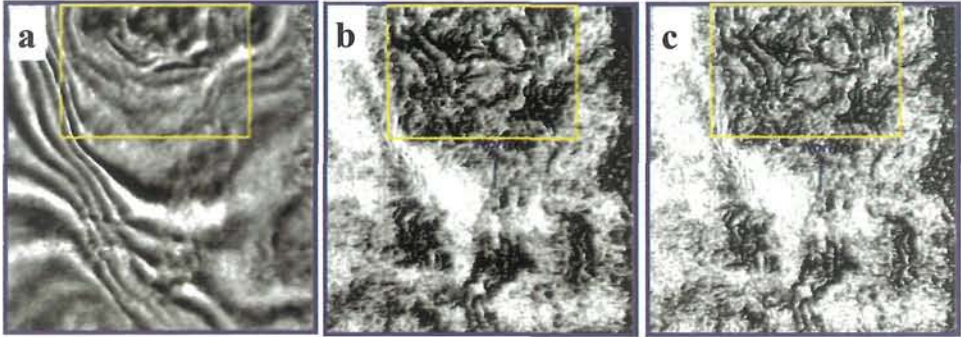


**Figure 11:** (a) Amplitude time slice of input data. It is a close-up of Figure 10A. (b) Time slice produced by applying edge detection to input data without EPS. It is a close-up of Figure 10B. (c) Time slice produced by applying edge detection to input data with EPS applied. It is a close-up of Figure 10C. (d) Interpretation of time slice in Figure 11C, which is produced by applying edge detection to input data with EPS applied. The channels are marked using blue lines and the faulting arrears are circled by red curves.

Besides being helpful for defining faults and channels, EPS also can help interpreters mapping fractures. It has been found that the edge-detection (coherence-cube) attribute is very helpful for mapping zones containing fractures and faults in Saudi Arabia (Lawrence, 1998). The area depicted by Figures 12 and 13 shows an area where fracture mapping is critical since the reservoir in this region is very tight (i.e., porosity of the

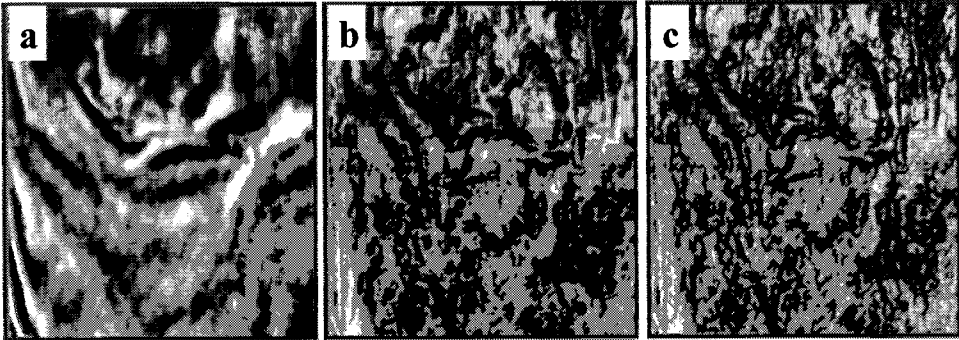
reservoir is less than 5%) and most of the hydrocarbon accumulates in highly fractured zones.

Figure 12A shows an amplitude time slice extracted from a 3-D seismic volume, Figure 12B displays the time slice of the edge-detection result without applying EPS, while 7C shows the edge-detection result with EPS applied before running the edge-detection algorithm. The edges in 7C are clearer and sharper than those in 7B.



**Figure 12:** (a) Amplitude time slice of input data. (b) Time slice produced by applying edge-detection algorithm to input data without EPS. (c) Time slice produced by applying edge-detection algorithm to input data with EPS.

Figure 13 is a close-up of Figure 12. The zooms confirm that the edge-detection results generated using data pre-processed with EPS are clearer and sharper than those using data without EPS applied. We have found that edge detection results enhanced by EPS (e.g., Figure 13C) allow us to see more details while interpreting fracture zones.

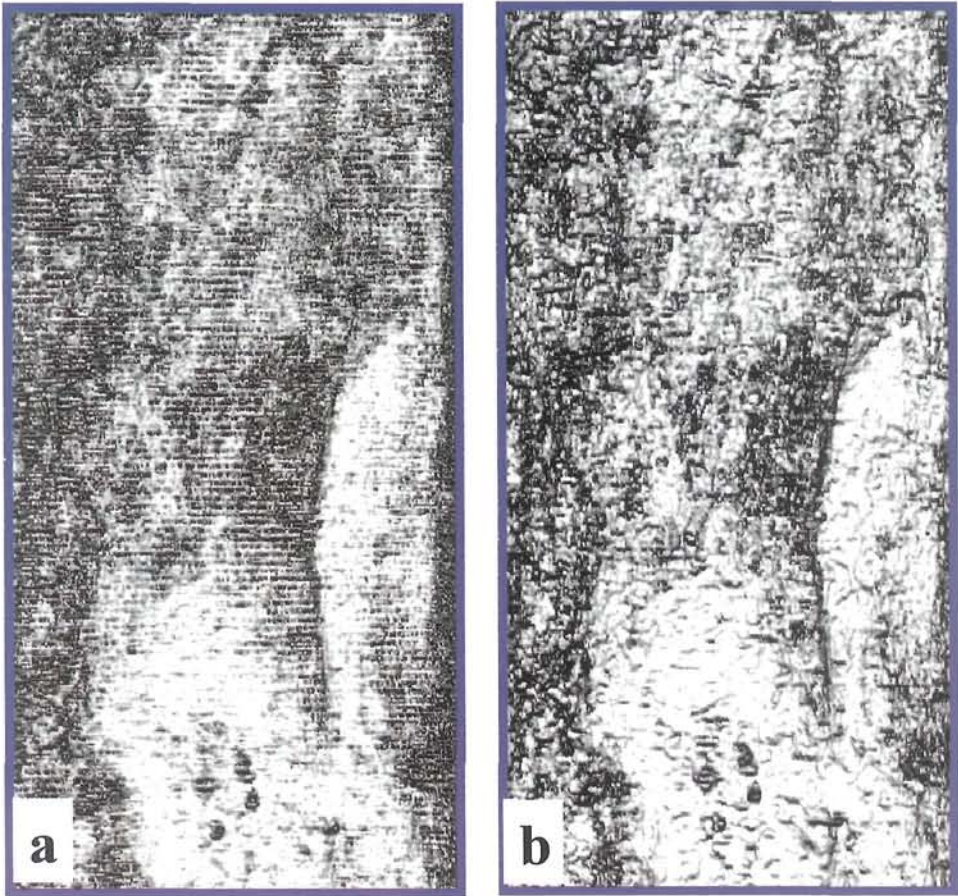


**Figure 13:** (a) Amplitude time slice of input data. It is a close-up of Figure 12A. (b) Time slice produced by applying edge detection to input data without EPS. It is a close-up of Figure 12B. (c) Time slice produced by applying edge detection to input data with EPS applied. It is a close-up of Figure 12C.

In addition to suppress random noise, the EPS can also be used to suppress certain type of coherence noise and acquisition artifacts. Figures 14A and 14B display the edge-detection results generated using field data without and with EPS applied, respectively. In 9A, strong acquisition footprints are clearly apparent, whereas they are barely visible in 9B after including EPS in the processing sequence. In this case, the footprints, which are not really random, can be suppressed by the EPS method because their width is smaller than the length of the EPS smoothing window in the cross-line direction.

In the experiment shown in Figure 14, we applied a 2-D square window on time slices in the 3-D seismic data volume. This square window is further divided into small wedge-shaped segments; the average over the segment with the minimum standard deviation is taken as the output at the center of the window. The window size used to get the result in Figure 14 is 5X5 samples. It is worth noting that such artificial footprints cannot be suppressed well if a smaller, say 3X3, window was used. That is, the noise, which may not be random in a strict definition, can be reduced as long as its character length is smaller than the size of the EPS window.





**Figure 14: Time slice produce by applying edge-detection algorithm to input data without EPS (a) and with EPS (b).**

In figure 14, the acquisition footprint can be suppressed by EPS method with a 5X5 smoothing window. This means that EPS can reduce the undesired noise as shown in Figure 14, on the other hand, genuine geological features (e.g., channels) would be suppressed if their width were smaller than the window size.

If small-size features are the desired output after running edge-detection, one should design a smaller EPS window (smaller than the character length of the expected features), or simply drop EPS from the processing sequence. Moreover, a new EPS method, which will not hurt non-random small features, has been developed and will be published in a separate paper as it is out of the scope of this paper.

#### 4 Conclusions

The EPS technique, if designed properly, can reduce random noise without altering genuine sharp boundaries that are related to geology. Unlike PEF or other similar methods, EPS does not heavily depend on the predictability of signals. For these reasons, the EPS method is an ideal pre-conditioning process for seismic edge-detection (or coherence-cube) algorithms.

#### 5 Acknowledgments

We thank P. G. Kelamis and K. P. Gunaratnam for reviewing and revising this manuscript.

#### References

1. Bahorich, M. S., and S. L. Farmer., 1995, 3-D seismic discontinuity for faults and stratigraphic features: The coherence cube, *The Leading Edge*, 1053-1058.
2. Caswell, F., 1995, *Success in STATISTICS*, Coventry University, 36-39.
3. Claerbout, J., 1998, *Geophysical exploration mapping*, Electronic book, Free software, Stanford Exploration Project (SEP) web site.
4. Gersztenkorn, A. and K. J. Marfurt, 1996, Eigen-structure based coherence computation, 66th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 328-331.
5. Laurence, P., 1998, Seismic attribute in Characterization of small-scale reservoir faults in Abqaiq field, *The Leading Edge*, 521-525.
6. Luo Y., M. AlFaraj, and M. I. Marhoon, 2002, Integration of New Techniques for Improving Interpretational Seismic Resolution, 72nd Ann. Internat. Mtg., Soc. Expl. Geophys., Salt Lake City, Expanded Abstracts, 2130-2133.
7. Luo Y., W. G. Higgs, and W. S. Kowalik, 1996, Edge detection and stratigraphic analysis using 3-D seismic data; 66th Ann. Internat. Mtg., Soc. Expl. Geophys. Expanded Abstracts, 324-327.
8. Marfurt, K., W. S. Duncan and P. Constance, 2002, Comparison of 3-D edge detection seismic attributes to Vinton Dome Louisiana, 72nd Ann. Internat. Mtg., Soc. Expl. Geophys., Salt Lake City, Expanded Abstracts, 577-580.
9. Nago, M and T. Matsuyama, 1980, *Complex Aerial photographs*, Kyoto University.

This page is intentionally left blank

## CHAPTER 21

### IMAGING SEISMIC REFLECTION DATA AT THE CORRECT DEPTH WITHOUT SPECIFYING AN ACCURATE VELOCITY MODEL: INITIAL NUMERICAL EXAMPLES OF AN INVERSE SCATTERING SUBSERIES

Simon A. Shaw

*BP Upstream Technology,  
200 Westlake Park Blvd.,  
Houston, TX 77079, U.S.A.  
E-mail: shawsa@bp.com*

Arthur B. Weglein

*Physics Department,  
and Geosciences Department,  
University of Houston,  
617 Science and Research Bldg. 1,  
Houston, TX 77204, U.S.A.  
E-mail: aweglein@uh.edu*

The objective of seismic depth imaging is to produce a spatially accurate map of the reflectivity below the Earth's surface. Current methods for performing depth imaging require accurate an accurate velocity model in order to place reflectors at their correct locations. Existing techniques to derive the wave velocity can fail to provide this information with the necessary degree of accuracy, especially in areas that are geologically complex.

The inverse scattering series, a multi-dimensional direct inverse procedure, has the potential to perform the task of imaging reflectors at depth without needing to specify the exact velocity. The primary objective of the research described here is to further define the concept and to progress the development of an algorithm to perform the task of imaging in the absence of accurate velocity information. As has been recently reported, the strategy employed involves isolating a subseries of the inverse series with the specific purpose of imaging reflectors in space.

In this paper, analytic and numerical results of an imaging subseries algorithm are further examined. This algorithm is being evaluated with regards to its convergence properties and data requirements.

#### 1. Introduction

Seismic reflection surveying is a very useful tool in the exploration and production of hydrocarbons. In the seismic experiment, a controlled source at the Earth's surface produces a wavefield that propagates in the Earth, reflects at subsurface interfaces, and is recorded by an array of receivers at the surface (Fig. 1). This experiment

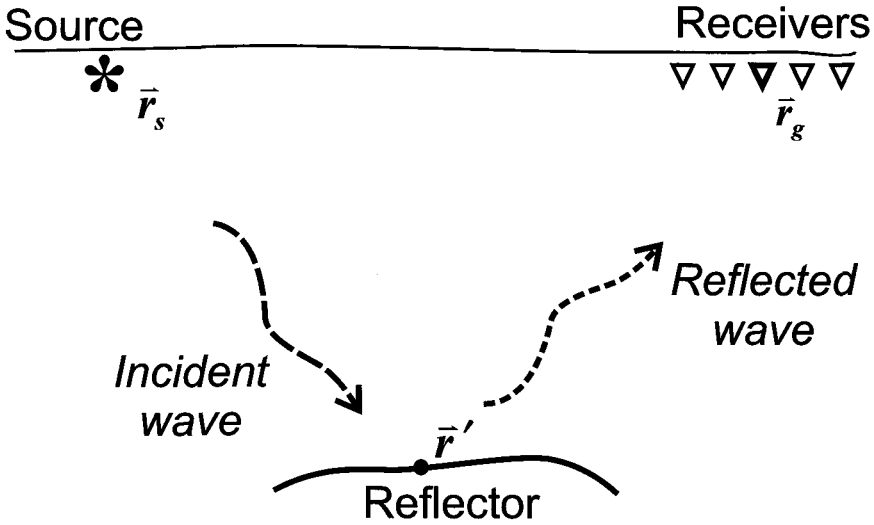


Fig. 1. Illustration of the seismic experiment. A seismic source sends an incident wave into the Earth. The wave reflects in the Earth and is recorded at one of many receivers. The recorded wavefield contains information about the source that created it, and the Earth through which it has traveled.

is repeated for many locations at the surface during the course of a survey. The seismic data measurements are processed to reveal information about the Earth's subsurface. These processed data are usually interpreted alongside regional geologic and other available geophysical data in the course of deciding where, and where not, to drill for hydrocarbon reserves.

At its core, seismic data processing is an inverse method: the data are inverted for the Earth's subsurface properties. These properties include the spatial location of reflectors and the contrasts in density and mechanical properties at these reflectors. In practice, processing of seismic data is carried out in a sequence of steps, e.g., random noise attenuation, source wavelet deconvolution, removal of free surface multiples, removal of internal multiples, imaging (also called migration), and inversion for changes in Earth properties. The order in which these steps are carried out can be important because most algorithms assume that the data have been preconditioned by the preceding processes. The research described here concerns the single step of imaging primaries. Imaging is the process which transforms the recorded primary seismic wavefield into a spatially accurate picture of the Earth's subsurface structure.

This paper represents progress in a long-term project to develop multi-dimensional algorithms that have an increased ability to achieve processing objectives while lowering the demands on (often inaccessible) a-priori information about the subsurface. As with earlier analysis of algorithms for multiple attenuation, the evaluation of new concepts and theory progresses from simple, analytic examples to

complex, numerical models and ultimately to field data. To avoid numerical, stability or discreteness issues, we begin testing with analytic data for one-dimensional examples. We are thereby able to isolate and characterize the results and attribute them to the inverse procedure only.

The following section explains the motivation for studying the particular problem of imaging when the exact velocity model is unknown. Then, in the next section, the inverse scattering series is derived and the strategy of isolating subseries of the inverse series that perform seismic processing tasks is described. Finally, in the remaining sections, analytic and numerical examples are used to demonstrate how a subseries of the inverse series is able to locate reflectors at their correct depths without specifying the velocity.

## **2. Motivation for an accurate imaging algorithm when the velocity is unknown**

Reflectors exist where there is a sharp contrast in Earth material properties. Reflectors are attributed to geological horizons or boundaries between different types of rocks and fluids. Oil and gas are often trapped below the surface by impermeable rocks. Seismic imaging produces a map of subsurface reflectors. The accuracy of this reflector map has a direct impact on our ability to predict the location, volume and even type of hydrocarbon reserves. Hence seismic imaging plays a key role in exploration and production of natural resources.

Traditional methods for imaging combine Green's Theorem with the wave equation to predict the wavefield inside the Earth from measurements on its surface<sup>1-4</sup>. This wavefield at depth is then transformed into a map of reflectivity using an imaging condition that asks for the seismic amplitude recorded in the limit of a small recording time for a hypothetical experiment where a source and receiver are coincident in the Earth.

These methods require the precise velocity model in order to compute the Green's functions that back-propagate the measured wavefield into the Earth. Without the true propagation velocity, the wavefield in the Earth will not be correctly predicted, and the imaging condition will fail to locate the reflectors. It is for this reason that the quality of the results from current methods for depth imaging are critically dependent on the accuracy of the velocity model.

Velocity information itself can be derived from seismic reflection data by picking reflection travel times<sup>5</sup> or using reflection tomography<sup>6</sup>. For current depth imaging algorithms, a velocity model is constructed consisting of interval velocity values that vary with subsurface location. The recorded wavefield is propagated through the velocity model to construct the wavefield in the Earth. In practice, the seismic interval velocities can be in error by 5-10%<sup>7</sup> depending on data quality, geologic complexity, and the sophistication of the algorithm being used to derive them.

The failure of current methods to produce accurate depth images below complex overburdens, such as below salt, basalt, and karsted or gas-saturated sediments, is

the motivation for the research described here. Under these geologic conditions, current velocity estimation procedures fail to yield the velocity model with the necessary degree of accuracy to place reflectors at their correct locations. The objective of this research is to develop a new method to accurately image seismic data that is less dependent on our ability to describe the precise velocity model.

### 3. Scattering theory and seismic data processing

The research described in this chapter applies inverse scattering theory to the seismic inverse problem<sup>8</sup>. In scattering theory, the difference in the behaviour of an incident wave in two media (referred to as the *reference* medium and the *actual* medium) is described in terms of the difference between the physical properties of these two media.

The wave equations for the actual and reference wavefields are expressed by

$$L\psi = A(\omega)\delta(\vec{r}_g - \vec{r}_s) \quad (1)$$

$$L_0\psi_0 = A(\omega)\delta(\vec{r}_g - \vec{r}_s) \quad (2)$$

where  $L$  and  $L_0$  are the differential operators that describe wave propagation in the actual and reference media, respectively,  $\psi$  is the actual wavefield,  $\psi_0$  is the reference wavefield, and  $A(\omega)$  is the source wavelet. The variables  $\vec{r}_g$  and  $\vec{r}_s$  are the receiver and source position vectors, respectively. The Green's functions for the reference medium and actual medium satisfy

$$L_0G_0 = \delta(\vec{r}_g - \vec{r}_s) \quad (3)$$

$$LG = \delta(\vec{r}_g - \vec{r}_s) \quad (4)$$

respectively, and so  $\psi = A(\omega)G$  and  $\psi_0 = A(\omega)G_0$ . The scattering potential and the scattered wavefield are defined by

$$V \equiv L_0 - L \quad (5)$$

$$\psi_s \equiv \psi - \psi_0 \quad (6)$$

respectively. The equation that relates the actual and reference wavefields to the scattering potential is the Lippmann-Schwinger equation:

$$\psi(\vec{r}_g|\vec{r}_s;\omega) = \psi_0(\vec{r}_g|\vec{r}_s;\omega) + \int_{-\infty}^{\infty} G_0(\vec{r}_g|\vec{r}';\omega)V(\vec{r}';\omega)\psi(\vec{r}'|\vec{r}_s;\omega)d\vec{r}'. \quad (7)$$

Equation 7 can be successively iterated for  $\psi$  in the integral. This results in the forward, or Born, series for the actual wavefield  $\psi$

$$\psi = \psi_0 + \psi_1 + \psi_2 + \dots \quad (8)$$

where

$$\psi_1(\vec{r}_g | \vec{r}_s; \omega) = \int_{-\infty}^{\infty} G_0(\vec{r}_g | \vec{r}'; \omega) V(\vec{r}'; \omega) \psi_0(\vec{r}' | \vec{r}_s; \omega) d\vec{r}' \quad (8a)$$

$$\psi_2(\vec{r}_g | \vec{r}_s; \omega) = \int_{-\infty}^{\infty} G_0(\vec{r}_g | \vec{r}'; \omega) V(\vec{r}'; \omega) \times \int_{-\infty}^{\infty} G_0(\vec{r}' | \vec{r}''; \omega) V(\vec{r}''; \omega) \psi_0(\vec{r}'' | \vec{r}_s; \omega) d\vec{r}' d\vec{r}'' \quad (8b)$$

⋮

The forward series is a solution for  $\psi$  in terms of  $G_0$ ,  $\psi_0$  and  $V$ . In other words, the wavefield that propagates in the *actual* medium is described in terms of an infinite series of propagations in a chosen *reference* medium and interactions with the potential  $V$ . Conversely, the inverse series is a solution for  $V$  in terms of the scattered field on the measurement surface  $(\psi - \psi_0)_m = (\psi_s)_m$  and  $G_0$ . The inverse series can be derived by first writing  $V$  as the sum of constituent components<sup>9</sup>

$$V = V_1 + V_2 + V_3 + \dots = \sum_{n=1}^{\infty} V_n \quad (9)$$

where  $V_n$  is the portion of  $V$  that is  $n^{\text{th}}$  order in the measured values of the scattered field,  $(\psi_s)_m$ . Substitution of equation 9 into the forward series (equation 8) and matching terms that are equal order in  $(\psi_s)_m$  yields the inverse series:

$$(G_0 V_1 \psi_0)_m = (\psi_s)_m \quad (9a)$$

$$(G_0 V_2 \psi_0)_m = -(G_0 V_1 G_0 V_1 \psi_0)_m \quad (9b)$$

$$(G_0 V_3 \psi_0)_m = -(G_0 V_1 G_0 V_1 G_0 V_1 \psi_0)_m - (G_0 V_2 G_0 V_1 \psi_0)_m - (G_0 V_1 G_0 V_2 \psi_0)_m \quad (9c)$$

⋮

See, e.g., Weglein et al. (1981)<sup>10</sup> for references to the development of the inverse series. To calculate only the first term in the inverse series (i.e., solving equation 9a) and to treat  $V_1 \approx V$  is to make the inverse Born approximation. However, it is important to note that the inverse series does not make that assumption.  $V_1$  is assumed to be the first order approximation to  $V$  and equation 9a is the exact equation for that quantity. The inverse Born approximation forms the basis of all current techniques employed to perform seismic inversion<sup>11</sup>, i.e., normal moveout (NMO) stack, amplitude variation with offset (AVO) analysis, migration (imaging) and migration-inversion<sup>12</sup>. Linear approximate inverse methods are also the basis of medical imaging and other non-destructive evaluation methods. For the seismic problem, the inverse Born is a reasonable approximation for precritical primary reflections, for small contrasts in material properties, and for a reference medium that is close to the actual medium. Second and higher terms in the inverse series can be viewed as correcting  $V_1$  towards  $V$  when the series converges. The tasks of removing multiples, imaging primaries at their correct depth, and inverting for



large changes in Earth properties reside in the second and higher order terms in the inverse series.

The inverse scattering series (equation 9) is a multi-dimensional direct inversion procedure. The scattering medium's properties are directly determined from the recorded data without iterative updating of the reference medium towards the actual medium. Alternative approaches, e.g., iterative linear inversion, involve updating the reference model so that the reference wavefield, in some sense, fits the observed data<sup>13</sup>. The inverse series is a distinct and separate method from iterative linear inversion. Equation 9 solves for  $V_1, V_2 \dots$  and hence for  $V = V_1 + V_2 + \dots$  directly in terms of  $(\psi_s)_m$  and  $G_0$ .

Empirical evaluation and tests of the *entire* inverse series have shown that it does not converge for contrasts between actual and reference medium properties greater than about 10 %<sup>14</sup>. Hence, it had been concluded that the radius of convergence was too small to be of direct practical use when no a-priori information is supplied. Rather than abandon the inverse series, research has been undertaken to isolate convergent subseries that perform individual tasks associated with inversion. Inversion of seismic data can be viewed as performing a sequence of four tasks:

- (1) Removal of free-surface multiples;
- (2) Removal of internal multiples;
- (3) Positioning of reflectors in space (imaging); and
- (4) Inverting reflectivity for changes in Earth parameters (target identification).

The inverse series accomplishes these tasks using only measured data and reference medium properties. Isolating specific subseries that perform these tasks is less ambitious than directly inverting for Earth properties in one step and so convergence properties may be more favorable. Also, by carrying out these tasks in sequence, tasks 2–4 benefit from the fact that previous tasks have already been performed, which constitutes valuable *a priori* information. At each step, the simplest possible reference medium is chosen that allows rapid convergence of the specific subseries.

This strategy first produced a multi-dimensional free surface multiple removal algorithm. Free surface multiples are events that have reflected in the subsurface, and then traveled back up, hit the free surface (at least once) and traveled back into the Earth. These events usually have large amplitudes (due to the high value of the reflection coefficient at the Earth's surface) and can obscure reflection events that have traveled further into the Earth but arrive at the same time as the multiples. Also, the presence of multiples often precludes accurate estimation of reference medium properties. The second task-specific subseries to be isolated was the one that predicts and attenuates internal multiple reflections. Internal multiples are events that have all their downgoing reflections below the free surface.

The multiple removal algorithms derived using the inverse series<sup>15</sup> have the unique property that they expect the recorded seismic data and the source wavelet as input, but do not require the propagation velocity or any other subsurface in-

formation. For marine seismic data, both the free surface and internal multiple subseries converge for a homogeneous acoustic reference medium – water – which makes the algorithms computationally efficient. More importantly, they predict and attenuate multiples generated by actual Earth model types that are much more complicated than the homogeneous acoustic reference medium and include elastic, heterogeneous, anisotropic and certain forms of anelastic media.

An important prerequisite of these inverse scattering algorithms is knowledge of the source wavelet. Methods for estimating the wavelet include direct near-field measurements<sup>16</sup> or an estimation from the recorded data<sup>17–19</sup>.

The strategy employed in this research is to first remove the multiples with their task-specific subseries, and then to use the source wavelet free demultiplied data as input to the subseries that act on primaries. This represents a staged approach where tasks are carried out in an order that progresses from easy to hard and that uses the successful completion of earlier tasks to improve the chances of the next more difficult task being successful.

Weglein et. al. (2000)<sup>20</sup> have proposed using the inverse scattering series to perform the third and fourth tasks of imaging reflectors in depth and inverting for Earth parameters, both in terms of reference medium information. The theory and concepts surrounding the task of imaging using the inverse series have been set out by Weglein et al. (2002)<sup>21</sup>. In this paper, the first numerical examples of an imaging subseries are presented and analyzed.

## 4. Imaging using the inverse series

### 4.1. 1-D inverse series and task separation

Wave propagation in a 1-D constant density variable velocity acoustic medium is described by the equation

$$\left( \frac{d^2}{dz^2} + k^2(z) \right) \psi(z; \omega) = 0 \quad (10)$$

where  $k(z) = \omega/c(z)$ ,  $\omega$  is the angular frequency,  $c(z)$  is the velocity, and  $z$  is the field point of the wavefield. Assume that the region that equation 10 describes does not contain the source. If the reference medium is chosen to be an acoustic wholespace with velocity  $c_0$ , then the perturbation has the form

$$\begin{aligned} V &= L_0 - L \\ &= k_0^2 - k^2(z) \\ &= k_0^2 \alpha(z) \end{aligned} \quad (11)$$

where  $\alpha(z) = (1 - c_0^2/c^2(z))$ . In this context, the inverse problem is to solve for  $\alpha$  where

$$\alpha = \alpha_1 + \alpha_2 + \alpha_3 + \dots = \sum_{n=1}^{\infty} \alpha_n \quad (12)$$

The first term in the inverse series (equation 9a) is then

$$\psi_s(z_m; \omega) = \int_{-\infty}^{\infty} \frac{e^{ik_0|z_m - z'|}}{2ik_0} k_0^2 \alpha_1(z') \psi_0(z'; k_0) dz' \tag{13}$$

where  $z_m$  is the measurement depth and the reference wavefield is  $\psi_0 = e^{ik_0(z' - z_m)}$ . Solving for  $\alpha_1$  yields

$$\alpha_1(z) = 4 \int_0^z \psi_s(z') dz' \tag{14a}$$

where  $z' = (c_0 t / 2 - z_m)$  and  $t$  is the travel time. Time zero ( $t = 0$ ) corresponds to when the downgoing incident wave passes  $z_m$ . Equations 9b and 9c become

$$\alpha_2(z) = -\frac{1}{2} \left( \alpha_1^2(z) + \left[ \frac{d\alpha_1(z)}{dz} \right] \int_0^z \alpha_1(z') dz' \right) \tag{14b}$$

$$\begin{aligned} \alpha_3(z) = & -\frac{1}{2} \left( 2\alpha_1\alpha_2(z) + \left[ \frac{d\alpha_1(z)}{dz} \right] \int_0^z \alpha_2(z') dz' \right. \\ & \left. + \left[ \frac{d\alpha_2(z)}{dz} \right] \int_0^z \alpha_1(z') dz' + \dots \right). \end{aligned} \tag{14c}$$

The equation for  $\alpha_2$  consists of two terms that correspond to “self-interaction” ( $\alpha_1^2$ ) and “separated” ( $\alpha_1' \int \alpha_1$ ) scattering diagrams as represented in Fig. 2. All higher order terms can be broken up in a similar manner. As has been reported<sup>21</sup>, separated diagram terms with a single “upward scattering point” contribute to a subseries that images reflectors at their correct spatial location, and self-interaction terms lead to the subseries that corrects the amplitude of  $\alpha_1$  towards  $\alpha$ . This will be illustrated in the next sections through analytic and numerical examples.

### 4.2. Analytic example of the imaging subseries

In this section, the ability of the inverse series to perform the task of imaging without needing to specify the velocity is illustrated using a simple 1-D acoustic example. Consider the experiment illustrated in Figure 3, with a source and receiver at the surface  $z_m = 0$ . The reference velocity is chosen to be constant  $c_0$ , whereas the actual Earth velocity is an unknown function  $c(z)$ .

In accordance with the strategy, all multiples (i.e., free-surface and internal multiples) have been removed from the input data<sup>14</sup>. Hence, for this example, the data consist of two primary reflections that arrive at times  $t_1$  and  $t_2$

$$\psi_s(t) = R_1 \delta(t - t_1) + \hat{R}_2 \delta(t - t_2). \tag{15}$$

$R_1$  is the reflection coefficient at the first interface for a downgoing wave and  $\hat{R}_2 = T_{01} R_2 T_{10}$  where  $R_2$  is the downgoing reflection coefficient at the second interface and  $T_{01}$  and  $T_{10}$  are the transmission coefficients for a wave propagating down and

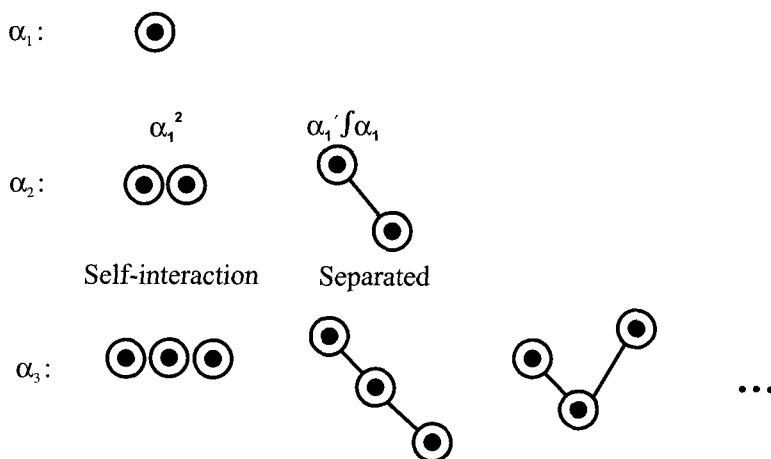


Fig. 2. The terms in the inverse series can be interpreted using scattering diagrams. The circles represent  $\alpha_1$  and the line represents a propagation in the reference medium. Self-interaction occurs when two or more scattering points are at the same location. Separated diagrams refer to scattering between points that are at different locations. Diagrams shown here perform inverse tasks on primary events.

up, respectively. Therefore

$$R_1 = \frac{c_1 - c_0}{c_1 + c_0} \quad (16)$$

$$T_{01} = 1 - R_1 \quad (17)$$

$$T_{10} = 1 + R_1 \quad (18)$$

$$\text{and } R_2 = \frac{c_2 - c_1}{c_2 + c_1} \quad (19)$$

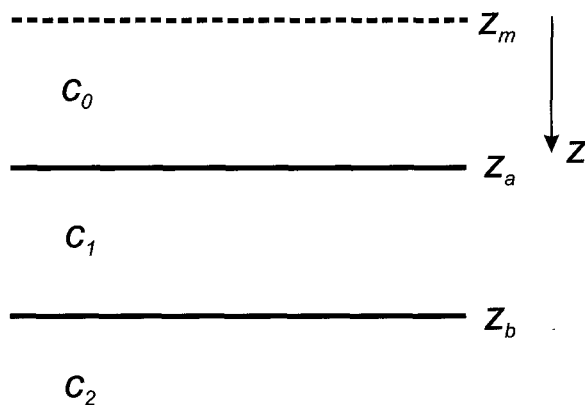


Fig. 3. A single layer with velocity  $c_1$  between two homogeneous half-spaces with velocity  $c_0$  and  $c_2$ . The depth of the first interface is  $z_a$  and the depth of the second interface is  $z_b$ .

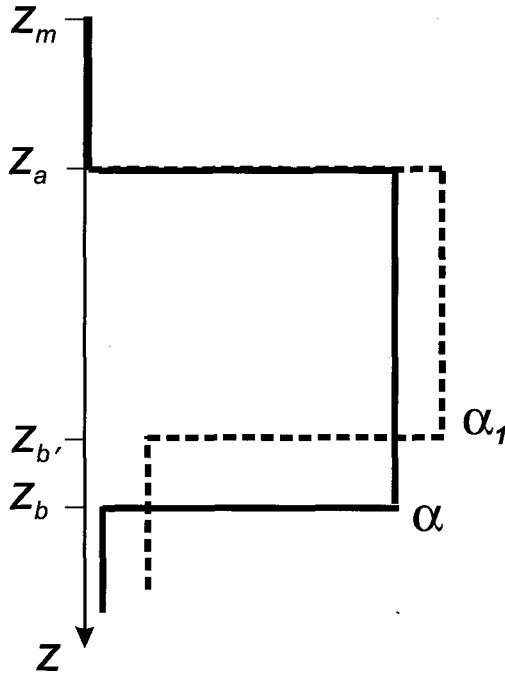


Fig. 4. Given the reference velocity  $c_0$ , the first term in the inverse series,  $\alpha_1$ , places the second interface at the incorrect depth  $z_{b'}$  rather than the actual depth  $z_b$ . For the case where  $c_0 < c_1$ ,  $z_{b'}$  will be shallower than  $z_b$ . The task of imaging is the process of moving the interface from  $z_{b'}$  to  $z_b$  whereas the last task in the inversion for  $\alpha$  must correct the amplitude of  $\alpha_1$ . Both of these tasks reside in the higher order terms of the inverse series.

Substitution of equation 15 into equation 14a yields

$$\alpha_1(z) = 4R_1H(z - z_a) + 4\hat{R}_2H(z - z_{b'}) \tag{20}$$

where  $z_{b'}$  is the pseudo depth at which the event with travel time  $t_2$  images with velocity  $c_0$ . This pseudo depth is equal to

$$z_{b'} = z_a + \gamma(z_b - z_a) \tag{21}$$

where  $\gamma = c_0/c_1$ . Figure 4 illustrates  $\alpha_1$  for the case where  $c_0 < c_1$ . The second reflector is imaged at a depth that is too shallow because the reference velocity is less than the actual velocity in the layer. Furthermore, the amplitude of  $\alpha_1$  is different from that of  $\alpha$ . These differences between  $\alpha_1$  and  $\alpha$  are corrected by the higher order terms in the inverse series.

Evaluating  $\alpha_2$  using equation 14b leads to

$$\begin{aligned} \alpha_2(z) = & -8 \left[ R_1^2H(z - z_a) + \hat{R}_2(2R_1 + \hat{R}_2)H(z - z_{b'}) \right] \\ & - 8 \left[ R_1\hat{R}_2(z - z_a)H(z - a)\delta(z - z_{b'}) \right] \end{aligned} \tag{22}$$

The terms consisting entirely of weighted Heaviside functions come from the self-interaction diagram (see Fig. 2) and act to correct the amplitude of  $\alpha_1$  towards  $\alpha$ . The  $\delta$  function comes from the separated diagram and is the first term in a series that shifts the deeper interface from  $z_{b'}$  to  $z_b$ . The imaging subseries algorithm accomplishes this shift through a Taylor Series for the difference of two Heaviside functions expanded about the mislocated interface. The Taylor Series for this shift may be written

$$\begin{aligned} B(z) &= H(z - z_{b'}) - H(z - z_b) \\ &= (z_b - z_{b'})\delta(z - z_{b'}) + \frac{(z_b - z_{b'})^2}{2}\delta'(z - z_{b'}) + \dots \end{aligned} \quad (23)$$

where we see that the first term in this series is a  $\delta$  function that relates to the one in equation 22. The coefficients of the Taylor Series contain the correct depth  $z_b$  to the interface. In the inverse series, these coefficients are constructed order by order in the measured values of the scattered field, and are a function of the amplitudes and travel times down to the reflector being imaged. For example, combining equations 16 and 21 provides the first coefficient

$$(z_b - z_{b'}) = 2(z_{b'} - z_a)(R_1 + R_1^2 + R_1^3 + \dots) \quad (24)$$

Convergence properties of the Taylor Series for a shift can be analyzed by performing a Fourier transform of  $B(z)$

$$\tilde{B}(k) = \int_{-\infty}^{\infty} \frac{e^{ik(z - z_{b'})}}{ik} \sum_{n=1}^{\infty} \frac{(ik(z_{b'} - z_b))^n}{n!} dk \quad (25)$$

and recognizing that the series

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad (26)$$

converges for any finite  $x$  where  $x = ik(z_{b'} - z_b)$  in equation 25. The quantity  $(z_{b'} - z_b)$  will be smaller when the reference velocity is closer to the actual velocity. From this Taylor Series analogy, it may be anticipated that the imaging subseries is convergent for any finite frequency, and for large contrasts between actual and reference medium velocity. Furthermore, from this analysis, we can expect that the rate of convergence will be greater for lower frequencies and when the perturbation is small. Hence, in practice, there will be a trade-off between rate of convergence and image resolution and the series will converge faster when the difference between the reference velocity and the actual velocity is small. Since the strategy is to remove multiples before processing primaries, then velocity analysis can be employed to derive a proximate velocity model.

In the next section, numerical examples illustrate how the imaging subseries terms act to shift the reflectors towards the correct depth for band-limited synthetic input data.

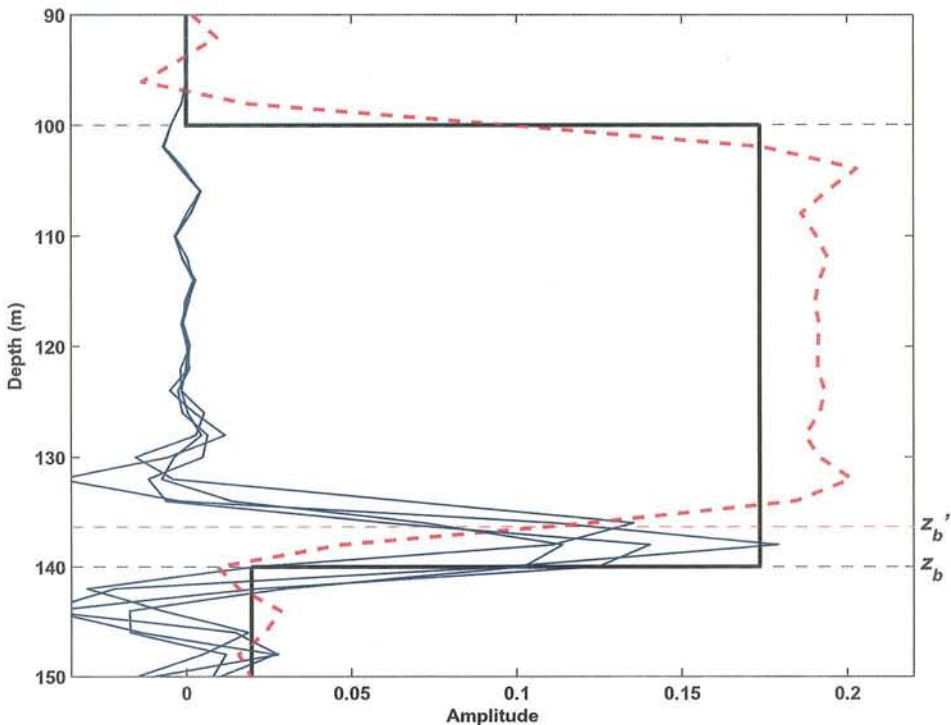


Fig. 5. Five terms in an imaging subseries. The solid black line is the actual perturbation  $\alpha$  and the dashed red line is  $\alpha_1$ , the first approximation to  $\alpha$ . The thinner blue lines are the non-linear imaging subseries terms. The cumulative sum of these imaging terms is shown in Figure 6.

## 5. Numerical Examples

Consider the 1-D model depicted in Fig. 3 with the following parameters:  $c_0 = 2000$  m/s,  $c_1 = 2200$  m/s,  $c_2 = 2020$  m/s,  $z_a = 100$  m and  $z_b = 140$  m. Choosing a reference velocity  $c_0 = 2000$  m/s, and simulating data for a 0 – 125 Hz band-limited source, then the computed  $\alpha_1$  is shown as the dashed red line in Fig. 5.

The depth that the reference velocity images the second reflector at is  $z_{b'} = 136$  m. The band-limited singular functions of the imaging subseries act to extend the interface from  $z_{b'}$  to  $z_b$ . The cumulative sum of these imaging subseries terms is illustrated in Fig. 6. After summing five terms, the imaging subseries has converged and the deeper reflector has moved towards its correct depth  $z_b$ .

Figure 7 shows the results of the imaging subseries for four possible velocity functions that are summarized in Table 1. In each case, the imaging subseries has imaged to the correct depth within seven terms. It has been found numerically that the subseries converges for large contrasts between actual and reference velocities. For example, when the actual velocity in the layer is 3000 m/s and the reference velocity is chosen to be 1500 m/s (i.e., a 100 % contrast) then the imaging series

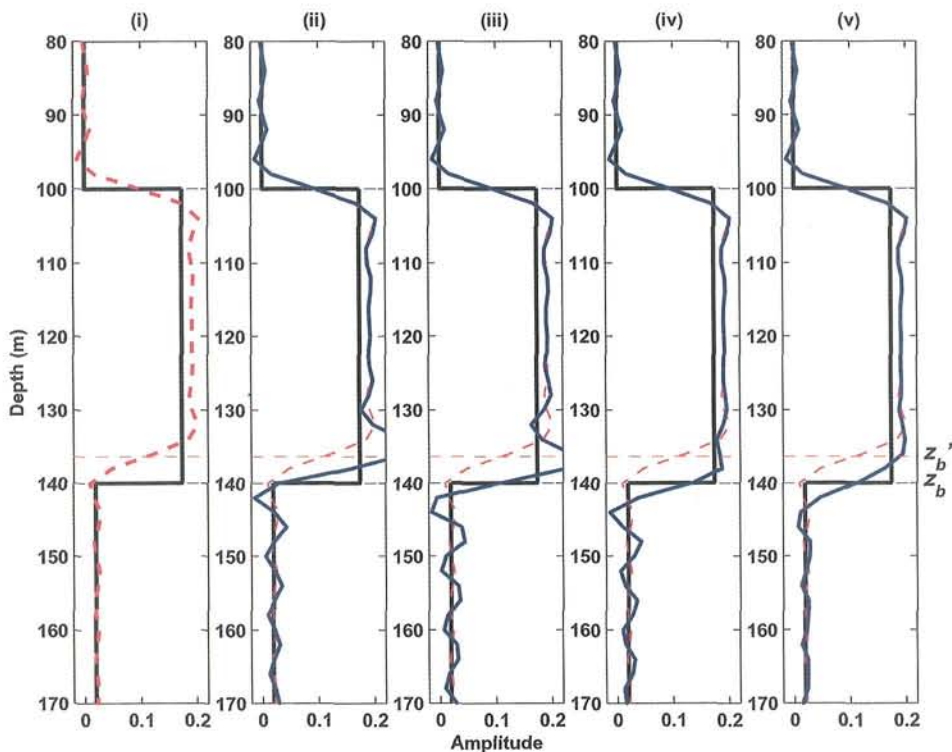


Fig. 6. Cumulative sum of five terms in an imaging subseries. The black solid line is the perturbation  $\alpha$  and the red line is the linear approximation to  $\alpha$  or the first term in the inverse series,  $\alpha_1$ . The blue line is the cumulative sum of the imaging subseries terms, e.g. in panel (ii) the sum of two terms in the subseries is shown, and in panel (v) the sum of five terms in the subseries is displayed.

Table 1. Velocity model parameters for the model illustrated in Fig. 3. The two reflectors are at  $z_a = 100\text{m}$  and  $z_b = 140\text{m}$ . Results of the imaging subseries algorithm in each case are shown in Fig. 7.

Figure	Velocity (m/s)		
	$c_0$	$c_1$	$c_2$
7(i)	2000	2200	2200
7(ii)	2000	1800	1980
7(iii)	2000	2100	2200
7(iv)	2000	1900	1800

will still converge. As expected, the number of terms required for convergence is found to be proportional to the contrast and the thickness of the layer.



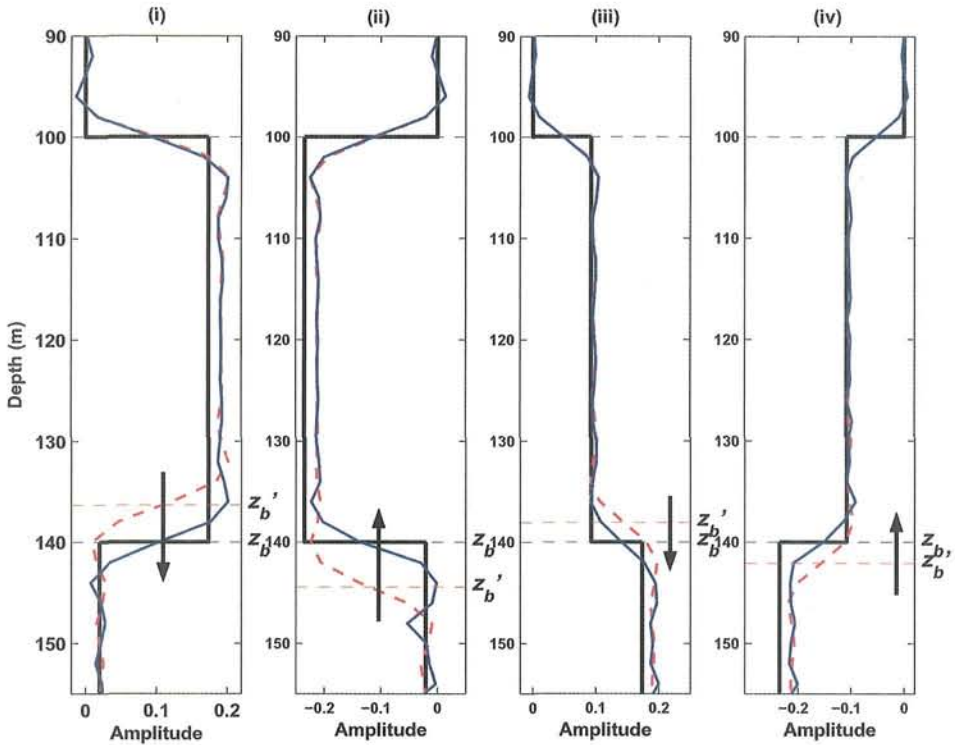


Fig. 7. Results of an imaging subseries for the velocity models summarized in Table 1. The solid black line is  $\alpha$  and the dashed red line is  $\alpha_1$ . The blue line is the cumulative sum of seven terms in the imaging series.

## 6. Conclusions

The inverse scattering series, a multi-dimensional direct inversion procedure, has the potential to achieve all the tasks normally associated with seismic inversion without requiring knowledge of the actual medium parameters. Recently, progress has been made in isolating a subseries of the inverse series that is responsible for imaging reflectors at their correct spatial location without knowledge of the actual velocity. This imaging subseries algorithm has been tested on 1-D analytic and synthetic data.

Further conceptual and algorithmic analyses, including the multi-dimensional extension, and generalizations for elastic wave propagation and spatially varying reference media, are the subjects of current research focus.

## Acknowledgments

We would like to thank Doug Foster, Ken Matson and Paulo Carvalho for contributions to the isolation of the subseries and early work investigating the use of the

inverse series to invert seismic data. Craig Cooper is thanked for his encouragement of this research. The support of the sponsors of the Mission-Oriented Seismic Research Program is gratefully acknowledged.

## References

1. J. F. Claerbout, *Geophysics*, Toward a unified theory of reflector mapping, **36**, 03, 467-481, (1971)
2. R. H. Stolt, Migration by Fourier transform, *Geophysics*, **43**, 01, 23-48, (1978)
3. W. A. Schneider, Integral formulation for migration in two-dimensions and three-dimensions, *Geophysics*, **43**, 01, 49-76, (1978)
4. C. P. A. Wapenaar and G. L. Peels and V. Budejicky and A. J. Berkhout, *Geophysics* Inverse extrapolation of primary seismic waves, **54**, 07, 853-863, (1989)
5. M. T. Taner and F. Koehler, Velocity spectra - Digital computer derivation and applications of velocity functions, *Geophysics*, **34**, 6, 859-881, (1969)
6. C. Stork and R. W. Clayton, Linear aspects of tomographic velocity analysis, *Geophysics* **56**, 4, 483-495, (1991).
7. S. H. Gray and J. Etgen and J. Dellinger and D. Whitmore, Seismic migration problems and solutions, *Geophysics* **66**, 5, 1622-1640, (2001).
8. A. B. Weglein, The inverse scattering concept and its seismic application, in *Developments in geophysical exploration methods*, A.A. Fitch (Ed.), Elsevier Applied Science, 06, 111-138, (1985)
9. H. Moses, Calculation of scattering potential from reflection coefficients *Phys. Rev.*, **102**, 559-567, (1956)
10. A. B. Weglein and W. E. Boyce and J. E. Anderson, *Geophysics*, Obtaining three-dimensional velocity information directly from reflection seismic data: An inverse scattering formalism, **46**, 08, 1116-1120, (1981)
11. R. W. Clayton and R. H. Stolt, A Born-WKB inversion method for acoustic reflection data, *Geophysics* **46**, 11, 1559-1567 (1981).
12. R. H. Stolt and A. B. Weglein, Migration and inversion of seismic data, *Geophysics* **50**, 12, 2458-2472 (1985).
13. A. Tarantola, *Inverse problem theory*, Elsevier Science B.V., (1987)
14. P. M. Carvalho, Free-surface multiple reflection elimination method based on non-linear inversion of seismic data, *Ph.D. Dissertation* (in Portuguese), *Universidade Federal da Bahia*, (1992)
15. A. B. Weglein and F. A. Gasparotto and P. M. Carvalho and R. H. Stolt, An inverse-scattering series method for attenuating multiples in seismic reflection data, *Geophysics*, **62**, 06, 1975-1989, (1997)
16. A. Ziolkowski, Why don't we measure seismic signatures? *Geophysics*, **56**, 2, 190-201, (1991)
17. A. Weglein and B. G. Secrest, Wavelet estimation for a multidimensional acoustic or elastic earth, *Geophysics*, **55**, 07, 902-913, (1990)
18. T. H. Tan, Wavelet spectrum estimation, *Geophysics*, **64**, 06, 1836-1846, (1999)
19. A. Osen, B. G. Secrest, L. Amundsen, and A. Reitan, Wavelet estimation from marine pressure measurements, *Geophysics*, **63**, 06, 2108-2119, (1998)
20. A.B. Weglein, K.H. Matson, D.J. Foster, P.M. Carvalho, D. Corrigan and S.A. Shaw, Imaging and inversion at depth without a velocity model: Theory, concepts and initial evaluation, *70th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts*, 1016-1019, (2000)
21. A.B. Weglein, K.H. Matson, D.J. Foster, S.A. Shaw, P.M. Carvalho and D. Corrigan,

Predicting the correct spatial location of reflectors without knowing or determining the precise medium and wave velocity: initial concept, algorithm and analytic and numerical example, *Journal of Seismic Exploration*, **10**, 367-382, (2002)

# **Time Series Prediction**

This page is intentionally left blank

## CHAPTER 22

### SEASONALITY EXTRACTION FROM TIME-SERIES OF SATELLITE SENSOR DATA

<sup>1,2</sup>Per Jönsson and <sup>3</sup>Lars Eklundh

<sup>1</sup>*Division of Mathematics, Natural Sciences and Language  
Malmö University, Malmö, Sweden*

<sup>2</sup>*Department of Physics, Lund University, Lund, Sweden  
E-mail: Per.Jonsson@ts.mah.se*

<sup>3</sup>*Department of Physical Geography and Ecosystems Analysis  
Lund University, Lund, Sweden  
E-mail: Lars.Eklundh@nateko.lu.se*

Two methods for processing time-series of satellite sensor data are presented. The first method is based on an adaptive Savitzky-Golay filter, and the second on non-linear least-squares fits to asymmetric Gaussian model functions. Both methods incorporate qualitative information on cloud contamination from ancillary datasets. The resulting smooth curves are used for extracting phenological parameters related to the growing seasons. The methods are applied to NASA/NOAA Pathfinder AVHRR Land Normalized Difference Vegetation Index (NDVI) data over Africa giving spatially coherent images of phenological parameters such as beginnings and ends of growing seasons, seasonally integrated NDVI, seasonal amplitudes etc. The results indicate that the two methods complement each other and that they may be suitable in different areas depending on the behavior of the NDVI signal.

#### 1. Introduction

To extract seasonality information it is necessary to generate smooth time-series from noisy satellite sensor data. This can be achieved by applying simple filters or by function fitting. A commonly used filtering method, BISE<sup>1</sup>, is easy to implement, but makes the usually erroneous assumption that all noise is negatively biased. Methods based on e.g. Fourier series<sup>2,3</sup> or on least-squares fits to sinusoidal functions<sup>4-6</sup> perform well if the shape of the time-profile is characterized by a well-defined annual cycle of growth and decline, but may fail when the time-profile is more ambiguous. Although most vegetation-covered areas are characterized by a clear annual growth and decline pattern, the shape of the time-profile may vary significantly between different bio-climatic zones. In some areas the time-profile is simple, allowing for the fitting of sinusoidal or bell-shaped functions. In other areas, particularly arid areas where the growing season is very short, the time-profile

has quite a different shape, and simple functions are often unable to describe the data. For this reason it is necessary to consider methods that are based on different assumptions regarding the shape of the annual time-profile. We present and test two new methods; adaptive Savitzky-Golay filtering with no prior assumptions on the time-profile, and non-linear least-squares fits to asymmetric Gaussians with the rather weak assumption that the growing season is characterized by a well-defined increase up to a peak level from which it declines until its end.

## 2. Data

The new methods are tested with the NASA/NOAA Pathfinder AVHRR Land (PAL) 8 km by 8 km database<sup>7,8</sup>. These data are well documented and have been used for global and regional studies for a number of years. Normalized Difference Vegetation Index (NDVI)<sup>9</sup>, rather than single-channel data, are used due to the observed relationships between the index and vegetation productivity<sup>10–13</sup>. Data are 10-day (decadal) maximum-value composites to reduce the effects of cloudiness<sup>14</sup>. Most remaining noise in the time-series is negatively biased, however, some positively biased noise occurs, e.g. due to anisotropic effects, sensor problems etc<sup>15</sup>. Although efforts have been made to calibrate PAL data for sensor differences, geometry and atmospheric disturbances, quality problems remain<sup>16</sup>. The CLAVR cloud flag channel is used as a simple indicator of data quality. In CLAVR universal thresholds in all the five AVHRR channels are used to classify pixels as clear, mixed and cloudy<sup>17</sup>. CLAVR data has been shown to underestimate clear pixels<sup>18</sup>. Despite this deficiency the information can be used to improve NDVI estimates<sup>19</sup>.

## 3. Methodology

We start by a general description of least-squares fits to an upper envelope. This is followed by an account on how to determine the number of annual seasons together with the approximate positions of their maxima and minima. The details of the adaptive Savitzky-Golay filter and the non-linear fitting to asymmetric Gaussians are given, and finally the extraction of seasonality information is discussed. The presented methods are implemented in a Fortran 90 program named TIMESAT and we refer to Jönsson and Eklundh<sup>20,21</sup> for computational aspects.

### 3.1. Least-squares fits to an upper envelope

Suppose that we have a time-series  $(t_i, I_i)$ ,  $i = 1, 2, \dots, N$  and a model function  $f(t)$  of the form

$$f(t) = c_1\varphi_1(t) + c_2\varphi_2(t) + \dots + c_M\varphi_M(t), \quad (1)$$

where  $\varphi_1(t), \varphi_2(t), \dots, \varphi_M(t)$  are given basis functions. Then the best values, in the least-squares sense, of the parameters  $c_1, c_2, \dots, c_M$  are obtained as the solution of

the system of normal equations

$$\mathbf{A}^T \mathbf{A} \mathbf{c} = \mathbf{A}^T \mathbf{b}, \quad (2)$$

where

$$A_{ij} = \frac{\varphi_j(t_i)}{\sigma_i}, \quad b_i = \frac{I_i}{\sigma_i}. \quad (3)$$

Here,  $\sigma_i$  is the measurement uncertainty of the  $i$ th data point. If these are not known they may all be set to the constant value  $\sigma = 1$ . To estimate the uncertainty of the NDVI data points, the PAL cloud flag channel (CLAVR) can be used. There are no general rules as for how the cloud information should be transformed into uncertainty estimates and judicious settings are up to the user. For the runs over Africa presented in this paper the uncertainty parameters are set to 1.0, 1.5 or 100.0 for data values corresponding to the CLAVR classes clear, mixed and cloudy. To take into account the fact that most noise, even for data classified as clear by CLAVR, is negatively biased, the determination of the parameters  $c_1, c_2, \dots, c_N$  of the model function is done in two steps<sup>20,21</sup>. In the first step the parameters are obtained by solving the system of normal equations with  $\sigma_i$  obtained from the ancillary data. Data points above the model function of the first fit are thought of as being more important, and in the second step the system is solved with the  $\sigma_i$  of the high data points decreased by some factor. The multi-step procedure leads to a model function that is adapted to the upper envelope of the data (Figure 1).

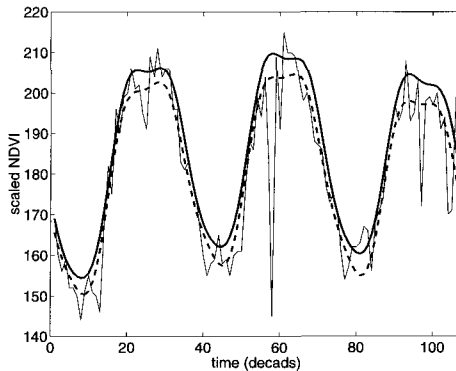


Fig. 1. Fitted functions from a two-step procedure. The dashed line shows a fitted function from the first step, and the solid line the fit from the second step.

### 3.2. Determination of the number of seasons

The high level of noise often makes it difficult to determine the number of annual seasons based on data for only one year. Including data from surrounding years



reduces the risk for erroneous determinations dramatically. In this work data values  $(t_i, I_i)$ ,  $i = 1, 2, \dots, N$  for three years are fitted to a model function

$$f(t) = c_1 + c_2t + c_3t^2 + c_4 \sin(\omega t) + c_5 \cos(\omega t) + c_6 \sin(2\omega t) + c_7 \cos(2\omega t) + c_8 \sin(3\omega t) + c_9 \cos(3\omega t) \quad (4)$$

where  $\omega = 6\pi/N$ . With 10-day data used in this study  $N = 108$ . The first three basis functions determine the base level and the three-year trend whereas the three pairs of sine and cosine functions correspond to, respectively, one, two and three annual seasons. The fitting procedure always gives three primary maxima and minima. In addition, secondary and tertiary maxima and minima may be found. If the amplitude of the secondary maxima exceeds a certain fraction of the amplitude of the primary maxima, there are two annual seasons. If the amplitude of the secondary maxima is low the number of annual seasons is set to one. In Figure 2 (a) the primary maxima and minima dominate and the number of seasons is set to one. In Figure 2 (b) the secondary maxima and minima are comparatively large and the number of annual seasons is set to two.

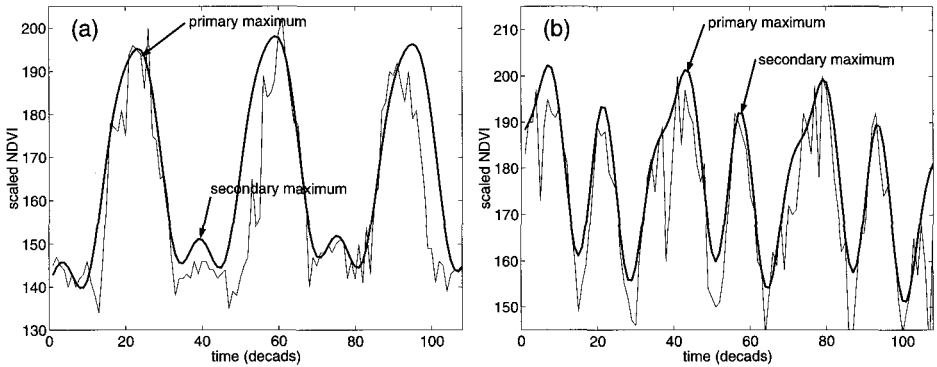


Fig. 2. Fits of sinusoidal functions and second order polynomials to three years of data in regions dominated by (a) one annual season and (b) two annual seasons. The thin solid line represents the original NDVI data. The thick solid line shows the fitted function.

### 3.3. Locally adapted Savitzky-Golay filtering

Savitzky-Golay filters are based on local polynomial fits<sup>22</sup>. For each point  $i = 1, 2, \dots, N$  a quadratic polynomial

$$f(t) = c_1 + c_2t + c_3t^2 \quad (5)$$

is fit to all  $2n + 1$  points in a window. The filtered value is then set to the polynomial value at this point. To account for the negatively biased noise, the fitting is done in multiple steps as described in the previous section. The result is a smoothed curve

that follows the upper envelope of the NDVI values. The width  $n$  of the moving window determines the degree of smoothing, but it also affects the ability to follow a rapid change. In TIMESAT two values of  $n$  can be set by the user. The first is used for data representing one annual season and the second for data representing two. Since, for large geographical areas, we are dealing with NDVI curves of different character it is desirable to use an adaptive method for  $n$ . Even if the global settings of the moving window work fairly well, it is sometimes necessary to locally tighten the window. A typical situation is in arid areas where the vegetation sometimes responds almost instantaneously to rainfall. To capture the corresponding sudden rise in data values, only a small window can be used. In TIMESAT the Savitzky-Golay filtering is performed using the global value  $n$  of the window. The filtered data are then scanned and if there is a large increase or decrease in an interval around a data point  $i$ , this data point will be associated with a smaller window. The filtering is then redone with the new locally adapted sizes of the window. Savitzky-Golay filtering with and without the adaptive procedure is illustrated in Figure 3.

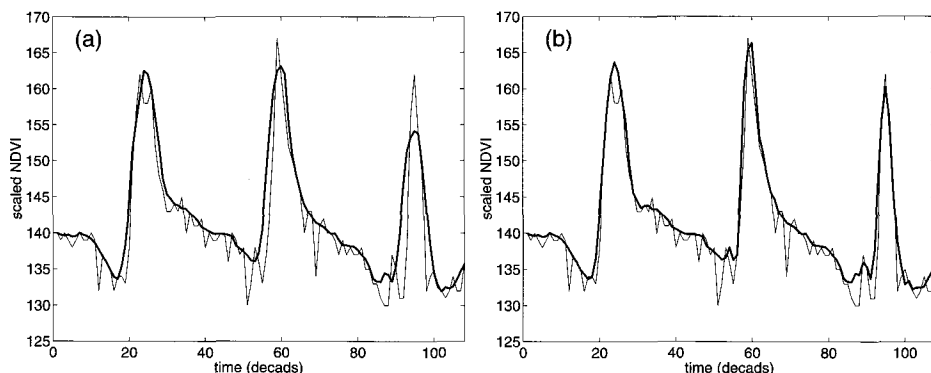


Fig. 3. Upper envelope Savitzky-Golay filtered data. Time is in ten day steps. In (a) the filtering is done with  $n = 5$ , which obviously is too large for the filtered data to follow the sudden increase and decrease of the underlying data values. A scan of the filtered data identifies the data points for which there is a large increase or decrease in surrounding intervals. Setting  $n = 3$  for these points and redoing the filtering gives the curve in (b). Note the improved fit at the rising edges and at the narrow seasonal peaks.

### 3.4. Least-squares fits to asymmetric Gaussian functions

In the asymmetric Gaussian method local model functions

$$f(t) \equiv f(t; c_1, c_2, a_1, \dots, a_5) = c_1 + c_2 g(t; a_1, \dots, a_5), \quad (6)$$

are fitted to data in intervals around maxima and minima. Here

$$g(t; a_1, \dots, a_5) = \begin{cases} \exp \left[ - \left( \frac{t-a_1}{a_2} \right)^{a_3} \right] & \text{if } t > a_1 \\ \exp \left[ - \left( \frac{a_1-t}{a_4} \right)^{a_5} \right] & \text{if } t < a_1 \end{cases} \quad (7)$$

is a Gaussian-type function. The linear parameters  $c_1$  and  $c_2$  determine the base level and the amplitude. For the Gaussian function,  $a_1$  determines the position of the maximum or minimum with respect to the independent time variable  $t$ , while  $a_2$  and  $a_3$  determine the width and flatness (kurtosis) of the right function half. Similarly,  $a_4$  and  $a_5$  determine the width and flatness of the left half. The local model functions are well suited for describing the shape of the scaled NDVI time-series in overlapping intervals around maxima and minima. Given a set of data points in an interval  $(t_i, I_i)$ ,  $i = n_1, \dots, n_2$  around a maximum or a minimum, the parameters  $c_1$ ,  $c_2$  and  $a_1, \dots, a_5$  are obtained by minimizing the merit function

$$\chi^2 = \sum_{i=n_1}^{n_2} \left[ \frac{f(t_i; c_1, c_2, a_1, \dots, a_5) - I_i}{\sigma_i} \right]^2. \quad (8)$$

The function depends non-linearly on the parameters  $a_1, \dots, a_5$  and in the program the minimization is done using an adaptive quasi-Newton method<sup>23</sup>. As in the previous cases the fitting is done in steps to account for the negatively biased noise. Given three local asymmetric Gaussian functions describing the left minimum, the central maximum and the right minimum (Figure 4 a), a global function fit can be built that describes the central season (Figure 4 b). The merging of local functions to a global function is a key feature of the method that increases the flexibility and allows the fitted function to follow a complex behavior of the time-series<sup>20</sup>.

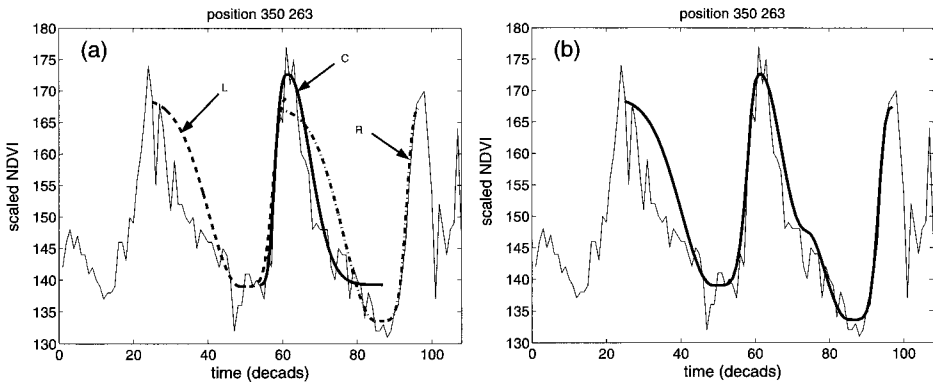


Fig. 4. (a) left (L), central (C) and right (R) local Gaussian functions. (b) merged global function. Note that the merged function in (b) only describes the vegetation of the central season from round decade 50 to decade 83. To describe vegetation of the left or right season additional local functions should be fitted and merged to global functions.

### 3.5. Extraction of seasonality parameters

Seasonal data are extracted for each of the growing seasons of the central year (Figure 5). Starting from the left base level or minimum, the beginning of a season is defined from the filtered or fitted functions as the point in time for which the value has risen by a certain amount, currently set to 10 % of the distance between the left base level and the maximum. The end of the season is defined in a similar way. The mid or the peak of a season is difficult to define, but a reasonable estimate is obtained as the position midway between the 90 % level positions above the left and right base levels. The annual integrated NDVI is frequently used in estimates of net primary production<sup>25–27</sup> through the relationship between NDVI and absorbed photosynthetically active radiation (APAR)<sup>12</sup>. To give a good estimate of the production of the phenologically dominant vegetation type it is also of interest to compute the integrated NDVI over the growing season, i.e. between the start and end of the season. In TIMESAT a small integral is defined as the area under the curve down to the mean of the left and right base levels. A large integral, extending to zero, is also defined.

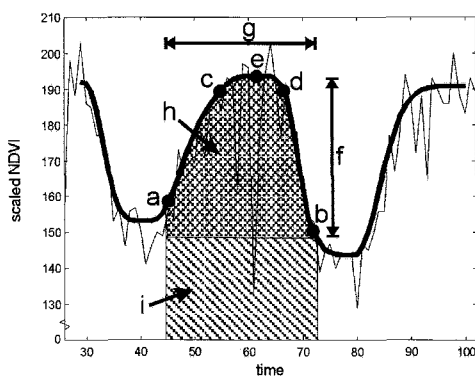


Fig. 5. Seasonality parameters computed in TIMESAT: (a) beginning of season, (b) end of season, (c) left 90 % level, (d) right 90 % level, (e) peak, (f) amplitude, (g) length of season, (h) integral over the growing season giving the area between the fitted function and the average of the left and right minimum values, (i) integral over the growing season giving the area between the fitted function and the zero level.

Other phenological parameters extracted are the peak values and the amplitude. The rate of increase in NDVI during the beginning of the season is theoretically related to the physiognomy of the vegetation and can be estimated by looking at the ratio between the amplitude and the time difference between the season start and the mid of the season. Another interesting quantity is the asymmetry, which can be defined as the ratio of the time differences between the mid of the season and the start and end of the season. A value of the asymmetry that is smaller than one indicate a rapid rise and a slow fall. Asymmetries larger than one, on the

other hand, are indicative of a slow rise and rapid fall. This type of behavior is often seen in areas with agricultural practices such as harvesting. Seasonal data can be extracted for single pixel locations and for whole images. Single pixel data are written to file and displayed with MATLAB. This is useful for testing the setting of parameters in specific locations. Image data are written to binary or ASCII files, and can be displayed with MATLAB or some suitable image processing software, like IDRISI, Easi/Pace or Erdas.

### 3.6. Results

We analyzed 1100 by 1060 pixels PAL image windows covering Africa over the period 1982-2000. For the sake of brevity, only a few examples are shown in this paper. These examples are selected to be representative of the differences between the two presented methods. Figure 6 shows the number of seasons, derived with the procedure outlined in Section 3.2. Areas characterized by two growing seasons (bi-modal) are mainly found in the Nile delta, parts of East Africa, along the Equator in central Africa, and in small areas along the coasts of West Africa. The sizes of these zones will vary somewhat with the parameter settings in the TIMESAT program. In Figure 6, pixels were classified as bi-modal if the amplitude of the secondary maxima were more than 40 % of the amplitude of the primary maxima. The observed pattern agrees well with what is expected given the general climatic circulation over Africa<sup>28</sup>.

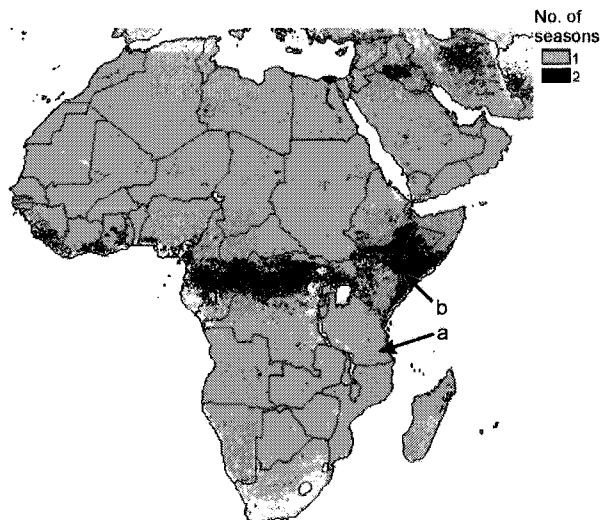


Fig. 6. Number of vegetational seasons in Africa 1999. Arrows point to uni-modal and bi-modal areas.

Figure 7 shows original NDVI and Savitzky-Golay filtered data for two areas, whose locations are indicated by the arrows in Figure 6. Figure 7 (a) is from a predominantly uni-modal (one season) area. Although there is a depression around the center of the central year, that might indicate the presence of two seasons, data is classified as uni-modal since the three years of data that have been used indicate that the area is dominated by a one seasonal cycle. Figure 7 (b), however, is bi-modal with two clearly developed growing cycles.

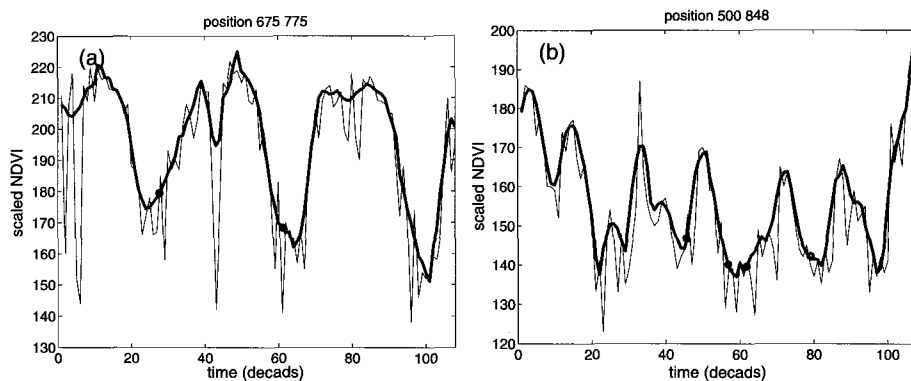


Fig. 7. NDVI and Savitzky-Golay filtered data from (a) uni-modal and (b) bi-modal areas, as indicated in Figure 7. Starts and ends of seasons are marked with circles.

Figure 8 displays original NDVI, Savitzky-Golay filtered data and asymmetric Gaussian functions for three pixels in Africa. The time-series in Figure 8 (a) is from the fringe of the Saharan desert, where the short annual season is dominated by a very rapid increase and decrease, followed by a slowly decreasing plateau. This composite behavior is typical for many extremely arid areas, and it is very well represented by the adaptive Savitzky-Golay method. However, the asymmetric Gaussian method has not represented the peak and the plateau of this time-series accurately. Figure 8 (b) is from a humid area with frequent cloudiness, resulting in a noisy time-series. The asymmetric Gaussian method has here generated a curve that is considerably smoother than the Savitzky-Golay filtered data. This smooth curve may be considered a better representative of the phenological curve of the vegetation, since some of the rapid changes seen in the Savitzky-Golay curve would not be expected to result from any physiological growth process, but rather from remaining noise due to clouds or other disturbances. Although the Savitzky-Golay curve follows the original data better, the smoothness of the asymmetric Gaussian curve might here be preferred. Figure 8 (c) was extracted from the Sahelian zone, well south of the Sahara. In this zone both the Savitzky-Golay and the asymmetric Gaussian function have modeled the time-series very well. Note that the beginnings and ends of seasons are located fairly close to each other in all three cases.

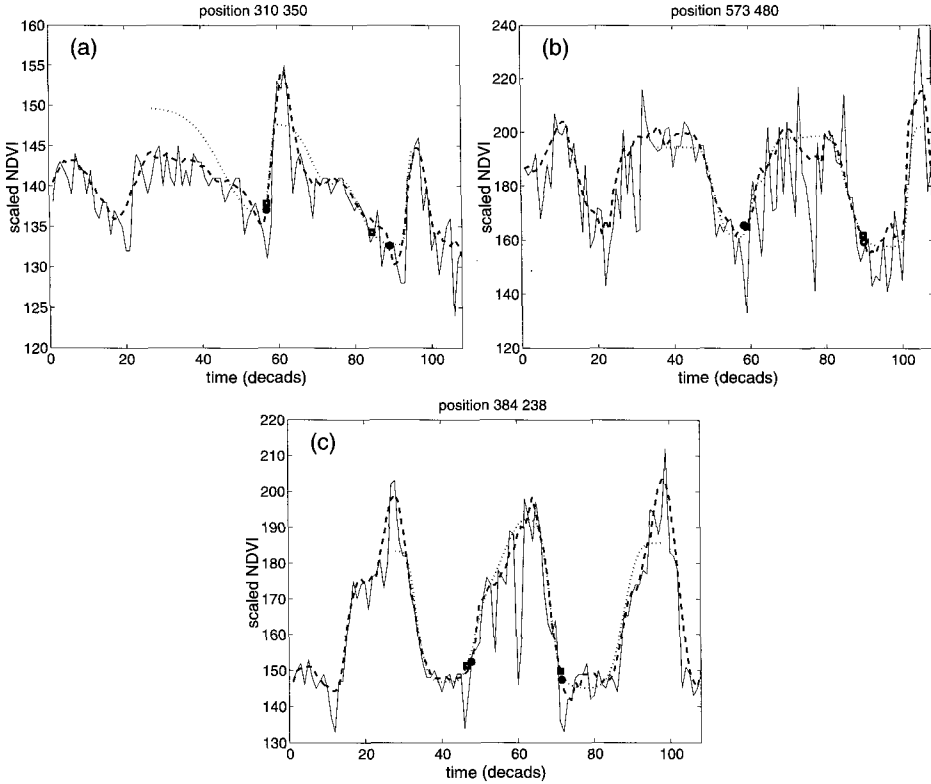


Fig. 8. Comparisons between Savitzky-Golay and asymmetric Gaussian curves for selected pixels. Beginnings and ends of seasons are marked by circles (Savitzky-Golay) and squares (asym. Gaussian). In (a) the Savitzky-Golay (dashed line) is superior to the asymmetric Gaussian (dotted line), which fails at following the rapid increase of the narrow central peak. In (b) the Gaussian method generates a smoother curve that better represents the behavior of vegetation than the curve generated by the Savitzky-Golay method. In (c) the both methods generate similar curves that both follow the data very well.

Figure 9 displays a number of parameters obtained with the Savitzky-Golay method. (a) gives the start of the first season in western Africa for 1999. Near the coast the season starts around decadal 5 (end of February). The starting date then shifts towards later dates until the border of the Sahara, where it falls at about decadal 25 (beginning of October). The observed pattern seems to be in general agreement with a climatic pattern dominated by the movements of the Inter Tropical Convergence Zone (ITCZ). In (b) the skewness or asymmetry is displayed. Note the belt with strong negative skewness that indicates a very rapid response to precipitation. (c) gives the peak value and (d) the amplitude for central Africa for the first growing season of 1999. In this area NDVI values are generally high. The seasonality of the evergreen vegetation, however, is not well developed and the resulting amplitudes are correspondingly small.

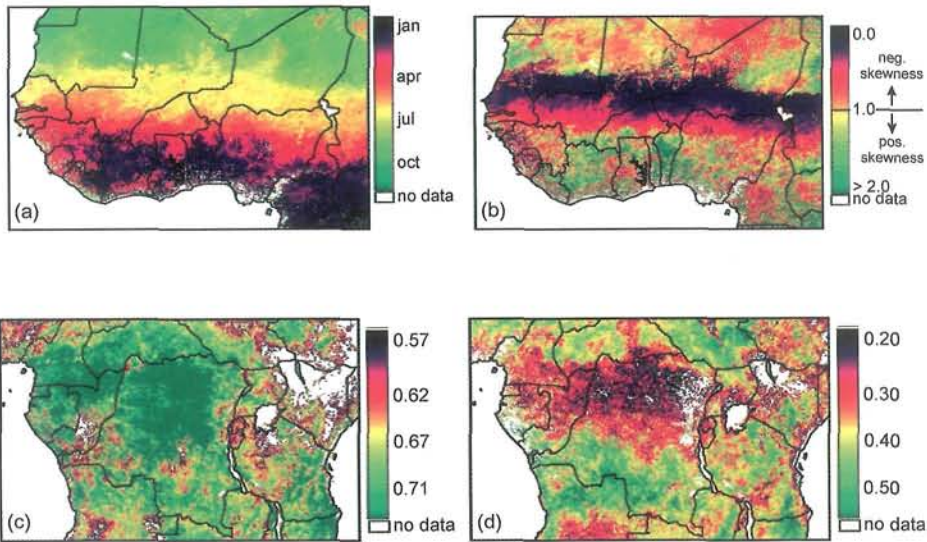


Fig. 9. Phenological parameters for the first growing season of 1999 extracted using Savitzky-Golay filtered data. (a) start of the growing season. (b) asymmetry or skewness. (c) peak value for the season. (d) amplitude of the season.

In order to investigate if there were any systematic differences between the two methods, values for some key phenological variables were extracted and averaged for each of the land-cover classes of the USGS digital land cover map<sup>29,30</sup>. Results for all classes, except those smaller than 5 % of the land area, are displayed in Table 1. The table indicates that the Savitzky-Golay method consistently generates higher values of amplitude and integrated NDVI than the asymmetric Gaussian method. Maximum values are also larger, but here the difference between the methods is smaller than for the other variables.

### 3.7. Discussion

Based on time-series of Pathfinder NDVI data the methods presented in this paper yield information about the seasonality of the underlying vegetation. Despite the high level of noise present in the original data the methods generate data that are spatially coherent and makes intuitive sense. AVHRR NDVI data are used in this study but other and newer sensor data are equally possible. The methods allow for ancillary data to be incorporated, and CLAVR cloudiness data is used as an indicator of uncertainties in the NDVI values. Although the effect of CLAVR was not explicitly tested, it is believed that new and better data quality indicators will improve the data fits. The seasonality patterns generated with the two



Table 1. Values for Savitzky-Golay (SG) and asymmetric Gaussian (AG) derived phenological variables computed as averages per USGS land cover class. Classes smaller than 5 % have been omitted.

Class	Percentage of pixels	SG ampl.	AG ampl.	SG max.	AG max.	SG large integral	AG large integral	SG small integral	AG small integral
Dryland cropland and pasture	5.9	45.2	39.2	187.3	186.1	5122.2	4681.2	821.7	709.9
Grassland	6.3	23.3	20.0	114.0	112.1	2991.0	2735.5	329.7	298.4
Shrubland	9.7	13.7	11.3	75.4	74.0	1483.7	1321.4	140.8	119.7
Savanna	25.8	47.4	41.3	191.5	190.4	5265.5	4841.1	843.5	740.4
Evergreen broadleaf forest	8.1	35.6	26.1	195.1	193.5	4853.9	4166.1	615.4	439.2
Barren or sparsely vegetated	32.4	0.6	0.5	3.3	3.2	85.3	75.8	9.9	8.3

methods largely correspond to each other with respect to the timings of seasonal events. However, some derived phenological parameters, notably the amplitude and the seasonal integrals, differ. The consistently lower values generated by the asymmetric Gaussian fits are believed to be underestimates due to the slower response to rapid phenological events. Although the timing of events is not much affected, amplitudes and integrals are. The possibility to modify the fits to handle these situations will have to be explored in future work. The asymmetric Gaussian functions generate smoother time-profiles that may represent the phenological behavior of vegetation better in some areas, notably when the time-series are very noisy. Thus, the two methods presented complement each other and the choice between them should be made based on the type of application and the behavior of the time-series in the specific area in mind. Other differences between the two methods are that the Savitzky-Golay algorithm is faster than the asymmetric Gaussian fit, and that it never fails to converge. This can happen with the non-linear Gaussian fit. However, of the 19 years of data over Africa analyzed, the asymmetric Gaussian fit only failed for a few hundred pixels, mostly over desert areas. The possibility to generate explicit information about the seasonality of the vegetation increases the potential use of time-series of satellite-derived spectral databases. It also underlines the importance of storing and maintaining long data series for the benefit of studies of phenological changes. Changes in key seasonality parameters might be used as early indicator of regional climatic changes.

### Acknowledgments

This work was supported by the Crafoord foundation. Lars Eklundh is employed on a grant from the Swedish National Space Board. Data used by the authors in this study include data produced through funding from the Earth Observing System Pathfinder Program of NASA's Mission to Planet Earth in cooperation with National Oceanic and Atmospheric Administration. The data were provided by Earth Observing System Data and Information System (EOSDIS), Distributed Active Archive Center at Goddard Space Flight Center which archives, manages, and distributes this data set.

### References

1. N. Viovy, O. Arino and A. S. Belward, *International Journal of Remote Sensing* **13**, 1585 (1992).
2. M. Menenti, S. Azzali, W. Verhoef and R. van Swol, *Advances in Space Research* **13**, 233 (1993).
3. L. Olsson and L. Eklundh, *International Journal of Remote Sensing* **15**, 3735 (1994).
4. J. Cihlar, *Remote Sensing of Environment* **56**, 149 (1996).
5. P. J. Sellers, C. J. Tucker, G. J. Collatz, S. O. Los, C. O. Justice, D. A. Dazlich and D. A. Randall, *International Journal of Remote Sensing* **15**, 3519 (1994).
6. G. J. Roerink, M. Menenti and W. Verhoef, *International Journal of Remote Sensing* **21**, 1911 (2000).

7. M. E. James and S. N. V. Kalluri, *International Journal of Remote Sensing* **15**, 3347 (1994).
8. J. R. G. Townshend, *International Journal of Remote Sensing* **15**, 3319 (1994).
9. J. W. J. Rouse, R. H. Haas, J. A. Schell and D. W. Deering, in *Third ERTS Symposium* (NASA SP-351, Washington DC, 1974), p. 309.
10. G. Asrar, M. Fuchs, E. T. Kanemasu and J. L. Hatfield, *Agronomy Journal* **76**, 300 (1984).
11. B. J. Choudhury, *Remote Sensing of Environment* **22**, 209 (1987).
12. C. J. Tucker and P. J. Sellers, *International Journal of Remote Sensing* **7**, 1395 (1986).
13. C. J. Tucker, C. Vanpraet, E. Boerwinkel and A. Gaston, *Remote Sensing of Environment* **13**, 461 (1983).
14. B. N. Holben, *International Journal of Remote Sensing* **7**, 1417 (1986).
15. G. G. Gutman, *Remote Sensing of Environment* **35**, 121 (1991).
16. S. D. Prince and S. N. Goward, *International Journal of Remote Sensing* **17**, 217 (1996).
17. L. L. Stowe, E. P. McClain, R. Carey, P. Pellegrino, G. Gutman, P. Davis, C. Long and S. Hart, *Advances in Space Research* **3**, 51 (1991).
18. S. Vemury, L. L. Stowe and V. R. Anne, *Journal of Atmospheric and Oceanic Technology* **18**, 169 (2001).
19. G. Gutman and A. Ignatov, *International Journal of Remote Sensing* **17**, 3295 (1996).
20. P. Jönsson and L. Eklundh, *IEEE Transactions on Geoscience and Remote Sensing* **40**, 1824 (2002).
21. P. Jönsson and L. Eklundh, *Computers and Geosciences*, (in preparation).
22. W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in FORTRAN. The art of scientific computing, Second edition*, (Cambridge University Press, Cambridge, 1992), p. 963.
23. J. E. Dennis, D. M. Gay and R. E. Welsch, *ACM Transactions on Mathematical Software* **7**, 348 (1981).
24. C. B. Field, M. J. Behrenfeld, J. T. Randerson and P. Falkowski, *Science* **281** 237 (1998).
25. A. Ruimy, B. Saugier and G. Dedieu, *Journal of Geophysical Research* **99**, 5263 (1994).
26. S. W. Running and R. R. Nemani, *Remote Sensing of Environment* **24**, 347 (1988).
27. S. N. Goward and D. G. Dye, *Advances in Space Research* **7**, 165 (1987).
28. H. E. Landsberg, *Climates of Africa. World Survey of Climatology*, (Elsevier, Amsterdam, 1972).
29. USGS/NASA, Africa Land Cover Characteristics Data Base Ver. 2.0. Internet reference: URL: [http://edcdaac.usgs.gov/glcc/afdoc2\\_0.html](http://edcdaac.usgs.gov/glcc/afdoc2_0.html). Accessed on 14 Aug. 2002, 2002.
30. T. R. Loveland, Z. Zhu, D. O. Ohlen, J. F. Brown, B. C. Reed and L. Yang, *Photogrammetric Engineering and Remote Sensing* **65**, 1021 (1999).

# **Image Compression**

This page is intentionally left blank

## CHAPTER 23

### NEAR-LOSSLESS COMPRESSION OF REMOTE-SENSING DATA

Bruno Aiazzi<sup>1</sup>, Luciano Alparone<sup>2</sup>, Stefano Baronti<sup>1</sup>, Cinzia Lastrì<sup>1</sup>

<sup>1</sup>IFAC-CNR: Institute of Applied Physics “N. Carrara”, National Research Council  
Via Panciatichi, 64, I-50127 Florence, Italy

E-mail: {aiazzi baronti}@ifac.cnr.it

<sup>2</sup>DET: Department of Electronics & Telecommunications, University of Florence  
Via Santa Marta, 3, I-50139 Florence, Italy

E-mail: alparone@det.unifi.it

This paper describes data compression algorithms capable to preserve the scientific quality of remote-sensing data, yet allowing a considerable bandwidth reduction to be achieved. Unlike *lossless* techniques, by which a moderate a compression ratio (CR) is attainable, due to intrinsic noisiness of the data, and conventional *lossy* techniques, in which the mean squared error of the decoded data is *globally* controlled by user, near-lossless methods are capable to *locally* constrain the maximum error, either absolute or relative, based on the user’s requirements. Advanced near-lossless methods rely on differential pulse code modulation (DPCM) schemes, based on either *prediction* or *interpolation*. The latter is recommended for lower quality compression (i.e., higher CR), the former for higher-quality, which is the primary concern in remote sensing applications. Experimental results of near-lossless compression of multispectral, hyperspectral, and microwave data from coherent imaging systems, like synthetic aperture radar (SAR), show the advantages of the proposed approach compared to standard lossy techniques.

#### 1. Quality Issues in Remote-Sensing Data Compression

Data compression is gaining an ever increasing relevance for the remote sensing community<sup>55</sup>. Since technological progresses allow observations of the Earth to be available at increasing spatial, spectral, radiometric, and temporal resolutions, the associated data volume is growing much faster than the transmission bandwidth does, either bandwidth of the downlink with the ground station for satellite platforms, or of the digital network supporting the data distribution. The introduction of data compression can alleviate bandwidth requirements at the price of a computational effort for encoding and decoding, as well as of a possible loss of quality.

Data compression consists of a decorrelation, aimed at generating a memoryless version of the correlated information source, followed by quantization, which introduces a distortion to yield a reduction in the information rate, and entropy coding<sup>38</sup>. If the decorrelation is achieved by means of an *orthonormal* transformation, e.g.,

the *discrete cosine transform* (DCT)<sup>46</sup>, or the *discrete wavelet transform* (DWT)<sup>39</sup>, the variance of quantization errors in the transformed domain is preserved when the data are transformed back to the spatial domain. Thus, the *mean square error* (MSE) can be easily controlled through the step sizes of quantizers. However, quantization errors in the transformed domain, which are likely to be uniformly distributed and are upper bounded in modulus by half of the step size, are spread by the inverse transformation and yield heavy-tailed distributions, whose maximum absolute amplitude cannot be generally known *a priori*. Therefore, lossy encoders, e.g., that proposed by the *Joint Photographic Experts Group* (JPEG)<sup>46</sup>, are unable to control the distortion but in the MSE sense, which means that relevant image features may be locally distorted by an unpredictable and unquantifiable extent<sup>54</sup>.

Compression methods can be either reversible, i.e., *lossless*, or irreversible (*lossy*), depending on whether images are exactly reconstructed after decoding or some *distortion* is introduced. A variety of image compression methods exists for applications in which reconstruction errors are tolerated. In remote sensing applications data modeling, features extraction, and classifications are usually performed<sup>24</sup>. Hence, the original quality of the data must often be thoroughly preserved after compression/decompression. As a matter of fact, however, the intrinsic noisiness of sensors prevents strictly lossless techniques from being used to obtain a considerable bandwidth reduction. In fact, whenever reversibility is recommended, compression ratios larger than two can hardly be obtained, because the attainable bit rate is lower bounded by the entropy of the sensor noise<sup>50,51,14</sup>.

Noteworthy are those lossy methods that allow to settle “a priori” the maximum reconstruction error, not only globally, but also locally. The maximum absolute error, also known as *peak error* (PE), or  $L_\infty$  distance between original and decoded image, is capable to guarantee a quality that is *uniform* throughout the image. If the  $L_\infty$  error is user defined, besides being constrained to be small, the current definition of *near-lossless* compression<sup>13</sup>, established for the medical community<sup>28</sup>, applies.

The evaluation of the maximum allowable distortion is an open problem. In the medical field objective measurements may be integrated with qualitative judgments of skilled experts, e.g., expressed in terms of *Receiver Operating Characteristic* (ROC) curves<sup>5</sup>. In remote sensing applications, however, photoanalysis is not the only concern<sup>45</sup>. The data are often postprocessed to extract information that may not be immediately available by user inspection. In this perspective, an attractive facility of near-lossless compression methods is that, if the  $L_\infty$  error is constrained to be, e.g., one half of the standard deviation of the background noise, assumed to be additive and independent of the signal, the decoded image will be *virtually lossless*<sup>9</sup>. This term indicates not only that the decoded image is visually indistinguishable from the original, but also that possible outcomes of postprocessing are likely to be practically the same as if they were calculated from the original data. Thus, the price of compression will be a small and predictable increment in the

equivalent sensor's noisiness<sup>12</sup>.

Focusing on images generated by coherent systems, like *synthetic aperture radar* (SAR) systems, lossless compression methods are little effective for reducing the data volume, because of the intrinsic noisiness which weakens the data correlation<sup>16</sup>. On the other hand, unpredictable local distortions introduced by lossy compression methods, though specifically tailored to SAR imagery<sup>32,63</sup>, may be unacceptable in many applications. Furthermore, the signal dependent nature of *speckle*<sup>35</sup> makes an  $L_\infty$  error control inadequate, since the errors to be bounded should not be *absolute* but *relative*, i.e., measured as pixel ratios, not as differences. In fact, larger errors should be encountered on homogeneous brighter areas than on darker ones. Hence, upper bounding of the PE no longer guarantees a quantifiable loss of quality. Therefore, near-lossless compression of SAR images should indicate that the pixel ratio of original to decoded image is strictly bounded within a prefixed interval<sup>17</sup>. If such an interval is comprised within the speckle distribution, then the decoded image will be *virtually lossless* as well<sup>12</sup>.

When multispectral or better hyperspectral data are being dealt with, *spectral* distortion becomes a primary concern, besides spatial and radiometric distortions. Spectral distortion is a measurement of how a pixel vector (i.e., a vector having as many components as spectral bands) changes because of an irreversible compression of its components. A widely used measurement is the angle between the two vectors. More sophisticated measurements based on information-theoretic criteria have recently proven themselves more effective in discriminating spectral classes<sup>27</sup>. For multiband data, whenever a virtually-lossless compression is unaffordable, a relative error-constrained compression may be rewarding in terms of scientific quality preservation of the decompressed data<sup>52</sup>. The rationale is that automatic analysis algorithms may be more sensitive to *relative* errors on pixels, than to *absolute* errors. For best performance, however, relative error-constrained compression requires logarithmic quantization<sup>15</sup>, which is penalized with respect to *linear* quantization in the *Rate Distortion* (RD) sense, with an MSE distortion measure<sup>38</sup>.

The problem of data transmission to ground stations is crucial for remote-sensing imaging systems orbiting on satellite platforms. Recent developments of advanced sensors originate huge amounts of data; however, once these data were lossy compressed, they would not be available as they were acquired for the user community. The Consultative Committee for Space Data Systems (CCSDS) has issued a recommendation for the lossless compression of space data<sup>29</sup>. Such a proposal has been recently adopted as an ISO standard<sup>36</sup>. Consequently, for what concerns on-board data compression, only lossless methods are presently recommended<sup>43</sup>. On the other side, to expedite dissemination and utilization of multispectral and especially hyperspectral images, near-lossless methods yielding constrained pixel error, either absolute or relative, are more suitable for obtaining a considerable bandwidth reduction and for preserving, at the same time, the spectral discrimination capability among pixel vectors, which is the principal source of spectral information.



## 2. Distortion Measures

### 2.1. Radiometric Distortion

Let  $\{g(i, j)\}$ , with  $0 \leq g(i, j) \leq g_{fs}$ , denote an  $N$ -pixel digital image and let  $\{\tilde{g}(i, j)\}$  be its possibly distorted version achieved by compressing  $\{g(i, j)\}$  and decompressing the outcome bit stream. Widely used distortion measurements are the following: Mean absolute error (MAE), or  $L_1$  (norm),

$$MAE = \frac{1}{N} \sum_i \sum_j |g(i, j) - \tilde{g}(i, j)|; \quad (1)$$

Mean Squared Error (MSE), or  $L_2^2$ ,

$$MSE = \frac{1}{N} \sum_i \sum_j [g(i, j) - \tilde{g}(i, j)]^2; \quad (2)$$

Root MSE (RMSE), or  $L_2$ ,

$$RMSE = \sqrt{MSE}; \quad (3)$$

Signal to Noise Ratio (SNR)

$$SNR_{(dB)} = 10 \cdot \log_{10} \frac{g^2}{MSE + \frac{1}{12}}; \quad (4)$$

Peak SNR (PSNR)

$$PSNR_{(dB)} = 10 \cdot \log_{10} \frac{g_{fs}^2}{MSE + \frac{1}{12}}; \quad (5)$$

Maximum absolute distortion (MAD), or *peak error*, or  $L_\infty$ ,

$$MAD = \max_{i,j} \{|g(i, j) - \tilde{g}(i, j)|\}; \quad (6)$$

Percentage maximum absolute distortion (PMAD)

$$PMAD = \max_{i,j} \left\{ \frac{|g(i, j) - \tilde{g}(i, j)|}{g(i, j)} \right\} \times 100. \quad (7)$$

Both in (4) and in (5) the MSE is incremented by the variance of the integer roundoff error, to handle the limit lossless case, when  $MSE = 0$ . Thus, SNR and PSNR will be upper bounded by  $10 \cdot \log_{10}(12 \cdot \bar{g}^2)$  and  $10 \cdot \log_{10}(12 \cdot g_{fs}^2)$ , respectively.

When multiband data are concerned, let  $v_l \triangleq g_l(i, j)$ ,  $l = 1, \dots, L$ , denote the  $l$ th component of the original multispectral pixel vector  $\mathbf{v}$  and  $\tilde{v}_l \triangleq \tilde{g}_l(i, j)$ ,  $l = 1, \dots, L$ , its distorted version. Some of the radiometric distortion measurements (1)-(7) may be extended to vector data as:

Average RMSE (A-RMSE), or  $L_1(L_2)$  (the innermost norm refers to vector space ( $l$ ), the outer one to pixel space ( $i, j$ )),

$$ARMSE = \frac{1}{N} \sum_{i,j} \sqrt{\sum_l [g_l(i, j) - \tilde{g}_l(i, j)]^2}; \quad (8)$$

Peak RMSE (P-RMSE), or  $L_\infty(L_2)$ ,

$$PRMSE = \max_{i,j} \sqrt{\sum_l [g_l(i,j) - \tilde{g}_l(i,j)]^2}; \quad (9)$$

$$SNR = 10 \cdot \log_{10} \frac{\sum_{i,j,l} g_l^2(i,j)}{\sum_{i,j,l} [g_l(i,j) - \tilde{g}_l(i,j)]^2}; \quad (10)$$

$$PSNR = 10 \cdot \log_{10} \frac{N \cdot L \cdot g_{fs}^2}{\sum_{i,j,l} [g_l(i,j) - \tilde{g}_l(i,j)]^2}; \quad (11)$$

Three-dimensional MAD, or  $L_\infty(L_\infty)$ ,

$$MAD = \max_{i,j,l} \{|g_l(i,j) - \tilde{g}_l(i,j)|\}; \quad (12)$$

Three-dimensional PMAD

$$PMAD = \max_{i,j,l} \left\{ \frac{|g_l(i,j) - \tilde{g}_l(i,j)|}{g_l(i,j)} \right\} \times 100. \quad (13)$$

In practice, A-RMSE (8) and P-RMSE (9) are respectively the average and maximum of the Euclidean norm of the distortion vector. SNR (10) is the extension of (4) to the 3-D data set. PSNR is the maximum SNR, given the full-scale of vector components. MAD (12) is the maximum over the set of pixel vectors of the maximum absolute component of the distortion vector. PMAD (13) is the maximum over the set of pixel vectors of the maximum percentage error over vector components.

## 2.2. Spectral Distortion

Given two spectral vectors  $\mathbf{v}$  and  $\tilde{\mathbf{v}}$  both having  $L$  components, let  $\mathbf{v} = \{v_1, v_2, \dots, v_L\}$  be the original spectral pixel vector  $v_l = g_l(i, j)$  and  $\tilde{\mathbf{v}} = \{\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_L\}$  its distorted version obtained after lossy compression and decompression, i.e.,  $\tilde{v}_l = \tilde{g}_l(i, j)$ . Analogously to the *radiometric* distortion measurements, *spectral* distortion measurement may be defined.

The spectral angle mapper (SAM) denotes the absolute value of the spectral angle between the couple of vectors:

$$SAM(\mathbf{v}, \tilde{\mathbf{v}}) \triangleq \arccos \left( \frac{\langle \mathbf{v}, \tilde{\mathbf{v}} \rangle}{\|\mathbf{v}\|_2 \cdot \|\tilde{\mathbf{v}}\|_2} \right) \quad (14)$$

in which  $\langle \cdot, \cdot \rangle$  stands for scalar product. SAM can be measured in either degrees or radians.

Another measurement especially suitable for hyperspectral data (i.e., for data with large number of components) is the spectral information divergence (SID)<sup>27</sup> derived from information-theoretic concepts:

$$SID(\mathbf{v}, \tilde{\mathbf{v}}) = D(\mathbf{v} \parallel \tilde{\mathbf{v}}) + D(\tilde{\mathbf{v}} \parallel \mathbf{v}) \quad (15)$$

with  $D(\mathbf{v}||\tilde{\mathbf{v}})$  being the Kullback-Leibler distance (KLD), or entropic divergence, or *discrimination*<sup>38</sup>, defined as

$$D(\mathbf{v}||\tilde{\mathbf{v}}) \triangleq \sum_{l=1}^L p_l \log \left( \frac{p_l}{q_l} \right) \quad (16)$$

in which

$$p_l \triangleq \frac{v_l}{\|\mathbf{v}\|_1} \quad \text{and} \quad q_l \triangleq \frac{\tilde{v}_l}{\|\tilde{\mathbf{v}}\|_1} \quad (17)$$

In practice SID is equal to the symmetric KLD and can be compactly written as

$$SID(\mathbf{v}, \tilde{\mathbf{v}}) = \sum_{l=1}^L (p_l - q_l) \log \left( \frac{p_l}{q_l} \right) \quad (18)$$

which turns out to be symmetric, as one can easily verify. It can be proven as well that SID is always nonnegative, being zero iff.  $p_l \equiv q_l$ ,  $\forall l$ , i.e., if  $\mathbf{v}$  is parallel to  $\tilde{\mathbf{v}}$ . The measure unit of SID depends on the base of logarithm: *nat/vector* with natural logarithms and *bit/vector* with logarithms in base two.

Both SAM (14) and SID (18) may be either averaged on pixel vectors, or the maximum may be taken instead, as more representative of spectral quality. It is noteworthy that radiometric distortion does not necessarily imply spectral distortion. Conversely, spectral distortion is always accompanied by a radiometric distortion, that is minimal when the couple of vectors have either the same Euclidean length ( $L_2$ ) for SAM, or the same city-block length ( $L_1$ ), for SID.

### 3. Compression of Raster Data

The term *raster data* denotes digitized samples of signals produced by sensors capable to detect and measure a physical property defined in a multidimensional domain. Examples of such signals are audio signals (1-D), scanned monochrome images (2-D), video signals (3-D), and generally most of remote-sensing data, including those produced by multispectral scanners and imaging spectrometers. A common feature of all raster data is *noisiness*, autocorrelation along each of their dimensions and a generally large number of digitization levels (at least 256, but usually more). Conversely, other data types, like ASCII texts, executable program codes, and other non-raster data files are characterized by a low number of symbols (usually  $< 256$ ), absence of *observation* noise and statistical dependencies among symbols resulting in the presence of repetitive patterns.

The classical lossless compression scheme consists of an optional decorrelator, followed by an entropy coding stage. A near-lossless scheme is similar and differs only for the presence of a quantizer. Hence, lossless compression can be regarded as a limit case in which the quantization step size  $\Delta$  is equal to 1. The decorrelator has the purpose of removing redundancy (when applicable, e.g., for raster data), and must hence be tailored to the specific characteristics of the data to be compressed. Examples are reversible integer transforms, e.g., integer wavelets<sup>25,47</sup>, and

causal/noncausal prediction<sup>46</sup>. Once a signal has been decorrelated, it is necessary to find a compact representation of its coefficients, possibly after they have undergone a quantization. Therefore an entropy coding algorithm is selected, which maps such coefficients into codewords, aiming at minimizing the average codeword length.

### 3.1. Source Decorrelation

When the statistical characteristics of the data to be compressed are known in advance (e.g., multimedia images, radar backscatters, seismic signals, multi/hyperspectral data), entropy coding techniques are seldom used standalone, but rather applied after a decorrelation stage, able to capture and remove the redundancy inherent in the nature of the data.

A large number of decorrelation techniques have been proposed in the scientific literature for different kinds of data. In the following we describe linear fixed prediction. A linear prediction is a *finite impulse response* (FIR) filter which, given the last  $P$  (the prediction order) received signal samples, attempts to estimate the value of the next one. The prediction error (i.e., the difference between each sample and its predicted value) tends to exhibit a lower dynamic range than the original signal, and can be hence quantized with lower distortion for a given bit rate. At the decoder, after initialization, the prediction error is added to the predicted signal, thus obtaining the original signal. In order to cope with the losses due to quantization, prediction is usually made from the quantized signal, both at the encoder and at the decoder. The coefficients of the prediction filter are generally computed on the basis of the second-order signal statistics (autocorrelation); hence the predictor ability to exploit knowledge of the signal correlation. This compression technique is usually referred to as Differential Pulse Code Modulation (DPCM), and is used in several compression algorithms for speech, images, and video. With respect to other decorrelation techniques, prediction has the advantage of being very simple and flexible, yet effective.

### 3.2. Entropy Coding Algorithms

A number of entropy coders, or reversible source coders, have been proposed in the past, such as Huffman and arithmetic coding, which have been established as “de facto” standards for the entropy coding stage of compression algorithms. Nearly all of them, more or less explicitly, exploit the concept of “variable length encoding”. The rationale is that, since the symbols of the input alphabet are usually not equiprobable, one can asymptotically attain the minimum average codeword length by assigning shorter codewords to symbols with higher probability of occurrence in the data stream. Among such coders, some deserve being considered in more detail. In the following we briefly describe Huffman coding, Golomb-Rice coding, arithmetic coding, dictionary coding and context-coding, and point out their suitability as entropy coding stages for raster data.

### 3.2.1. *Huffman coding*

Huffman coding<sup>38</sup> is perhaps the simplest example of a variable-length coder. This technique operates by sorting the symbols of the input alphabet according to their probabilities of occurrence in the message. In a first phase a tree is built by combining, at each pass, the two least probable symbols; a zero is assigned to the least probable of the two, and a one to the other. Then the codewords are obtained by walking from the root of the tree to the leaves, and writing down the zeroes and ones encountered. Due to its near-optimality in the exploitation of the first-order statistics of the input alphabet, Huffman coding has gained much popularity, and has been used as the entropy coding stage of several compression standards (e.g., JPEG)<sup>46</sup>.

### 3.2.2. *Golomb-Rice coding*

Given two integers,  $l \geq 0$  and  $m \geq 2$ , Golomb coding<sup>34</sup> of  $l$  is the binary representation of  $l \bmod m$  (remainder of the integer division of  $l$  by  $m$ ) followed by the *unary* coding of  $\lfloor l/m \rfloor$ , i.e.,  $\lfloor l/m \rfloor$  zeroes followed by a one or vice-versa. Golomb coding is asymptotically optimum for geometrically distributed ergodic sequences<sup>42</sup>, where  $m$  depends on the variance of the distribution. Golomb-Rice (GR) codes<sup>48</sup> are a subclass of Golomb codes in which  $m = 2^k$ : hence  $l \bmod 2^k$  are the  $k$  LSBs of the binary representation of  $l$  and if  $n = \lceil \log_2(l) \rceil$ , then  $\lfloor l/2^k \rfloor$  are the  $n - k$  MSBs of  $l$ . GR codes, although suboptimal with respect to Golomb codes<sup>42</sup>, are suitable for space applications due to their simple structure. GR codes can be made adaptive by properly choosing  $k$ , e.g., based on the spatial context of prediction errors (remapped to positive integers). In the following GR codes will be referred to as *Rice codes*, as indicated by CCSDS<sup>30</sup>.

### 3.2.3. *Arithmetic coding*

Arithmetic coding<sup>59</sup> recasts the transmission of symbols into the representation of the intervals of the cumulative distribution associated to the probability of occurrence of the symbol. Each interval, and hence symbol, is specified by its extrema, and is coded by means of the number in the interval exhibiting the shortest representation in the desired basis, e.g., binary. Very likely symbols have the largest intervals, and large intervals are more likely to contain numbers with compact representations. Therefore, more frequent symbols will be represented by shorter codewords. Arithmetic coding is known to be closer than Huffman coding to the optimal performance, but its implementation may be critical, especially in a space environment.

### 3.2.4. *Dictionary coding*

Huffman, Golomb-Rice, and arithmetic encoders rely either on the knowledge of the probabilities of the data to be encoded, or equivalently on an assumed probability

model, e.g., Laplacian. Dictionary coding belongs to the class of *universal* coding algorithms, in which a symbol string is compressed without a priori knowledge on the distribution of the source that has generated the data. Unlike other variable-length coders, dictionary-based algorithms encode variable-length strings of symbols (phrases) as single tokens, or better fixed-length indexes into some dictionary. A *dictionary* can be defined as a list of phrases that are expected to occur most frequently. The well-known Lempel-Ziv algorithm,<sup>64</sup> which is successfully applied to text compression, is perhaps the most popular dictionary encoder. The Lempel-Ziv algorithm achieves asymptotically optimum compression for strings generated by any stationary ergodic source and, in many cases, it performs the task in a quite practical manner. Unfortunately the algorithm does not ensure good compression results when the alphabet of the original source is large, as in the case of lossless coding of raster data in general.

### 3.2.5. Context coding

Context coding<sup>49</sup> is another universal coding algorithm devised to encompass the limitations of dictionary coding. Consider a data string  $(x_1, x_2, \dots, x_n)$  generated by a source with memory. A context is assigned to each symbol  $x_i$ ,  $1 \leq i \leq n$  based on the past  $x_1, x_2, \dots, x_{i-1}$  symbols. A context is nothing else than an equivalence class selected on some relevant feature of the past sequence. Classifying the original data into contexts aims at skewing the empirical distribution given by symbol occurrence counts in order to lower the conditional entropy. The application of context coding to image compression<sup>57</sup> has launched the new generation of advanced image encoders<sup>26</sup>.

## 4. Advanced Near-Lossless Image Compression Algorithms

Considerable research efforts have been recently spent in the development of lossless image compression techniques. The first specific standard has been the lossless version of JPEG<sup>46</sup>, which may use either Huffman or arithmetic coding. More interestingly, a new standard, which provides also near-lossless compression, has been recently released under the name JPEG-LS<sup>58</sup>. It is based on an adaptive nonlinear prediction and exploits context modeling followed by Golomb-Rice entropy coding. A similar context-based algorithm named CALIC has also been recently proposed<sup>60</sup>. The simple adaptive predictors used by JPEG-LS and CALIC, however, the *median adaptive predictor* (MAP) and the *gradient adjusted predictor* (GAP), are empirical. Thorough comparisons with more advanced methods<sup>21</sup> have revealed that their performance is limited and still far from the entropy bounds. Eventually, it is worth mentioning that Part I of the JPEG2000 image coding standard<sup>37</sup> foresees a lossless mode, based on reversible integer wavelets, and capable to provide a scalable bit stream that can be decoded from the lossy (not near-lossless) up to the lossless level. However, image coding standards are not suitable for the compression

of 3-D data sets: in spite of their complexity, they are not capable to exploit the three-dimensional signal redundancy featured, e.g., by multi/hyperspectral imagery.

Besides a number of integer-to-integer transforms<sup>2,3,25,47,53</sup>, capable of ensuring a perfect reconstruction with integer arithmetics, DPCM schemes, either *causal* (prediction-based) or *noncausal*, i.e. interpolation-based or *hierarchical*<sup>4</sup>, are indeed the only algorithms suitable for lossless compression of 2-D data, or more generally for  $L_\infty$ -constrained (near-lossless) compression. It is noteworthy that, unlike a locally MMSE linear prediction, a nonlinear prediction, like GAP of CALIC and MAP of JPEG-LS, that may occur to minimize the Mean Absolute Error (MAE), does not ensure local entropy minimization<sup>40</sup>. Therefore only linear prediction, yet adaptive, will be concerned in the following for a 3-D extension suitable for multispectral and hyperspectral data.

DPCM basically consists of a decorrelation (see Sect. 3.1) followed by entropy coding of the outcome prediction errors. The simplest way to design a predictor, once a *causal* neighborhood is set, is to take a linear combination of the values of such a neighborhood, with coefficients optimized in order to yield *minimum mean squared error* (MMSE) over the whole image. Such a prediction, however, is optimum only for stationary signals. To overcome this drawback, two variations have been proposed: *adaptive* DPCM (ADPCM)<sup>46</sup>, in which the coefficients of predictors are continuously recalculated from the incoming new data, and *classified* DPCM<sup>33</sup>, in which a preliminary training phase is aimed at recognizing some statistical classes of pixels and at calculating an optimized predictor for each class. Such predictors are then adaptively combined<sup>31,8</sup> (as limit case the output is switched among one of them<sup>6,15</sup>), to attain the best space-varying prediction. This strategy will be referred to as adaptive combination/switching of adaptive predictors (ACAP/ASAP).

While details of up-to-date ASAP schemes, both 2-D<sup>22</sup> and 3-D<sup>15</sup> will be reviewed in Sect. 5, the ACAP paradigm underlies the development of a novel fuzzy logic-based prediction<sup>21</sup> (FMP) in which images are first partitioned into blocks and an MMSE linear predictor calculated for each block. From the large number of predictors obtained, a fuzzy-clustering algorithm produces an initial guess of a user specified number of prototype predictors to be fed to an iterative procedure in which to each predictor pixels are given degrees of membership measuring the fitness of prediction on another causal neighborhood larger than the prediction support; then predictors are recalculated from pixels depending on their degrees of membership. The overall prediction will be fuzzy, being given by the sum of the outputs of each predictor weighted by the memberships of the current pixel to that predictor. The linearity of prediction makes it possible to formulate the above approach as a problem of approximating the optimum space-varying linear predictor at each pixel by projecting it onto a set of nonorthogonal prototype predictors capable to embody the statistical properties of the image data.

The ACAP paradigm has been extended also to 3-D data<sup>16</sup>, same as ASAP<sup>15</sup>, by simply changing the 2-D neighborhood into a 3-D one spanning up to three

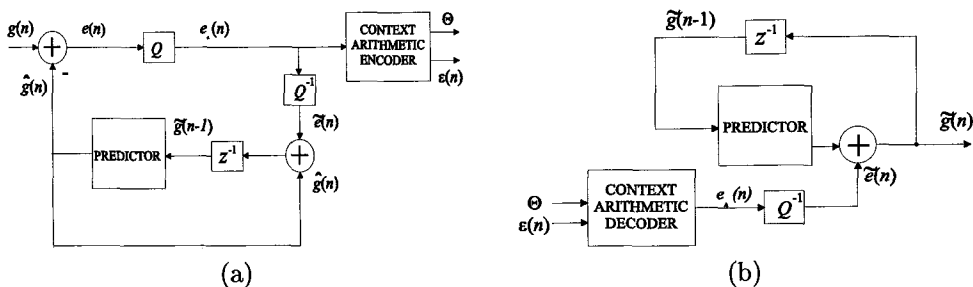


Fig. 1. Flowchart of DPCM with quantization noise feedback loop at the encoder, suitable for error-bounded near-lossless compression: (a) encoder; (b) decoder.

previous bands. To enhance the entropy coding performance, both schemes exploit context modeling (see Sect. 4.1) of prediction errors followed by arithmetic coding. It is noteworthy that the original 2-D FMP<sup>21</sup> achieves lossless compression ratios 5% better than CALIC and 10% than JPEG-LS, on an average. Although RLPE (2-D) is slightly less performing than FMP (2-D), its feature of real-time decoding is highly valuable in application contexts, since an image is usually encoded only once, but decoded many times. Details of this scheme can be found in Sect. 5.

#### 4.1. Context Modeling

A notable feature of all the advanced data compression methods<sup>22,21,26,31,53,58,60</sup> is statistical context modeling for entropy coding. The underlying rationale is that prediction errors should be similar to stationary white noise as much as possible. As a matter of fact, they are still spatially correlated to a certain extent and especially are non-stationary, which means that they exhibit space-varying statistics. The better the prediction, however, the more noise-like prediction errors will be.

Following a trend established in the literature, first in the medical field<sup>44</sup>, then for lossless coding in general<sup>53,60,58</sup>, and recently for *near-lossless* coding<sup>20,61</sup>, prediction errors are entropy coded by means of a classified implementation of an entropy coder, generally arithmetic or Golomb-Rice. For this purpose, they are arranged into a predefined number of statistically homogeneous classes based on their spatial *context*. If such classes are statistically discriminated, then the entropy of a *context-conditioned* model of prediction errors will be lower than that derived from a stationary memoryless model of the decorrelated source<sup>57</sup>.

A context function may be defined and measured on prediction errors lying within a causal neighborhood, possibly larger than the prediction support, as the RMS value of prediction errors (RMSPE). The context function should capture the nonstationary of prediction errors, regardless of their spatial correlation. Again, causality of neighborhood is necessary in order to make the same information available both at the encoder and at the decoder. At the former, the probability density function (PDF) of RMSPE is calculated and partitioned into a number of intervals chosen as equally populated; thus, contexts are equiprobable as well. This choice is



motivated by the use of adaptive arithmetic coding for encoding the errors belonging to each class. Adaptive entropy coding, in general, does not require previous knowledge of the statistics of the source, but benefits from a number of data large enough for training, which happens simultaneously with coding. The source given by each class is further split into sign bit and magnitude. The former is strictly random and is coded as it stands, the latter exhibits a reduced variance in each class, –null if the context (RMSPE) of the current pixel is always equal its magnitude; thus, it may be coded with few bits. It is noteworthy that such a context-coding procedure is independent of the particular method used to decorrelate the data. Unlike other schemes, e.g., CALIC<sup>60</sup>, in which context-coding is embedded in the decorrelation procedure, the above method<sup>20</sup> can be applied to any DPCM scheme, either lossless or near-lossless.

## 5. Near-lossless Compression Through 3-D Causal DPCM

Whenever multispectral images are to be compressed, advantage may be taken from the spectral correlation of the data for designing a prediction that is both *spatial* and *spectral*, from a causal neighborhood of pixels<sup>8,23,41,51,56,62</sup>. Causal means that only previously scanned pixels on the current and previously encoded bands may be utilized for predicting the current pixel value. This strategy is as more effective as the data are more spectrally correlated, as in the case of hyperspectral data<sup>8</sup>. If the *interband* correlation of the data is weak, as it usually occurs for data with few and sparse spectral bands, a 3-D prediction may lead to negligible coding benefits. In this case, advantage may be taken from a *bidirectional* spectral prediction<sup>10</sup>, in which once the  $(k - 1)$ st band is available, first the  $k$ th band is skipped and the  $(k + 1)$ st band is predicted from the  $(k - 1)$ st one; then, both these two bands are used to predict the  $k$ th band in a spatially causal but spectrally noncausal fashion.

The DPCM encoder utilized in this work is based on a classified linear-regression prediction according to the ASAP paradigm, followed by context-based arithmetic coding of the outcome residues. Image bands are partitioned into blocks, typically  $8 \times 8$ , and an MMSE linear predictor is calculated for each block. Given a prefixed number of classes, a clustering algorithm produces an initial guess of as many classified predictors that are fed to an iterative labeling procedure which classifies pixel blocks simultaneously refining the associated predictors.

In order to achieve reduction in bit rate within the constraint of a near-lossless compression<sup>1</sup>, prediction errors are quantized with odd valued step sizes,  $\Delta = 2E + 1$ , with  $E$  denoting the induced  $L_\infty$  error, with a quantization noise feedback loop embedded into the encoder, so that the current pixel prediction is formulated from the same “noisy” data that will be available at the decoder (see Fig. 1(a)).

For the case of a relative-error bounded compression a rational version of prediction error must be envisaged. Let us define the *relative prediction error* (RPE)

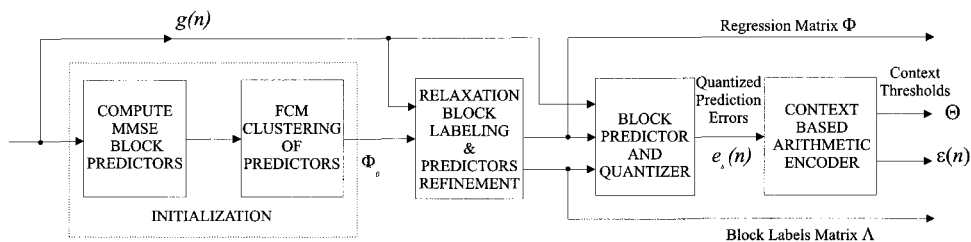


Fig. 2. Flowchart of the causal DPCM encoder with context coding referred to as *relaxation-labeled prediction encoder* (RLPE) in the following.

as ratio of original to predicted pixel value:

$$r(n) \triangleq \frac{g(n)}{\hat{g}(n)} \quad (19)$$

The *rational* nature of RPE, however, makes linear quantization unable to guarantee a strictly user-defined relative-error bounded performance.

Given a step size  $\Delta \in \mathbb{R}$ , with  $\Delta > 0$  and  $\Delta \neq 1$ , let us define as *logarithmic* quantization (Log-Q) of  $t \in \mathbb{R}$ ,  $t > 0$ ,

$$Q_{\Delta}(t) \triangleq \text{round}[\log_{\Delta}(t)] = \text{round}[\log(t)/\log(\Delta)] \quad (20)$$

$$Q_{\Delta}^{-1}(l) = \Delta^l$$

Applying (20) to (19) yields

$$Q_{\Delta}[r(n)] = \text{round}\left[\frac{\log(g(n)) - \log(\hat{g}(n))}{\log \Delta}\right] \quad (21)$$

If a Log-Q with a step size  $\Delta$  is utilized to encode pixel RPEs, it can be proven that the ratio of original to decoded pixel value is strictly bounded around one

$$\min\left\{\sqrt{\Delta}, \frac{1}{\sqrt{\Delta}}\right\} \leq \frac{g}{\hat{g}} \leq \max\left\{\sqrt{\Delta}, \frac{1}{\sqrt{\Delta}}\right\} \quad (22)$$

Now, let us introduce a peak measurement of rational error, namely the *peak rational error*, defined as

$$PRE_{(\text{dB})} \triangleq 20 \log_{10} \left( \frac{\max\left[\frac{g(n)}{\hat{g}(n)}\right]}{\min\left[\frac{g(n)}{\hat{g}(n)}\right]} \right) \quad (23)$$

It is easily verified that if Log-Q is utilized, then

$$PRE_{(\text{dB})} = 20 \cdot \log_{10}(\Delta) \quad (24)$$

thus, the peak rational error it may be easily user-defined.

Quantized prediction errors are then arranged into activity classes based on the spatial *context*, which are entropy coded by means of arithmetic coding. Fig. 2 shows the flowchart of the encoder. As it appears, the refined predictors are transmitted along with the label of each block and the set of thresholds defining the context classes for entropy coding.

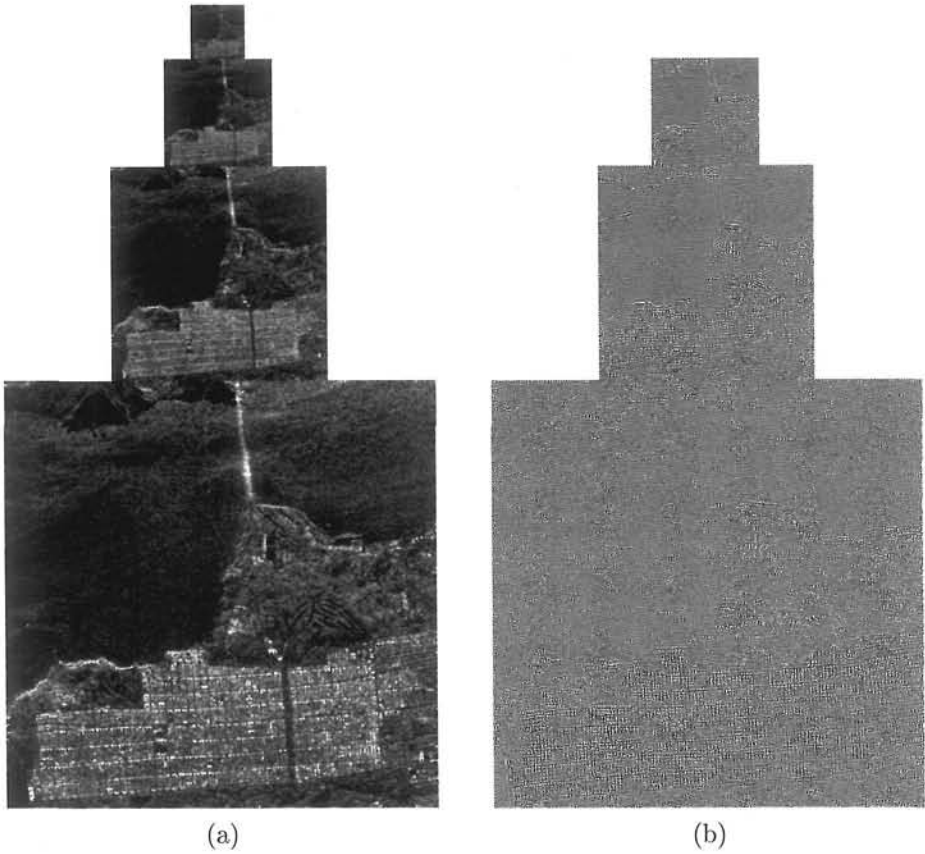


Fig. 3. (a) Gaussian pyramid and (b) rational Laplacian pyramid of test SAR image.

## 6. Near-lossless Image Compression Through Noncausal DPCM

Laplacian pyramids are multiresolution image representations obtained through recursive *reduction* (lowpass filtering followed by downsampling) and *expansion* (upsampling followed by lowpass filtering).

Start with  $G_0(i, j) \equiv g(i, j)$ ; define the Gaussian pyramid (GP) as

$$G_{k+1} = \text{reduce}_2\{G_k\} \quad k = 0, 1, \dots, K-1 \quad (25)$$

in which  $2^K$  is the largest power of two in which the image size can be factorized, and

$$\text{reduce}_2\{G_k\} \triangleq (G_k \otimes r_2) \downarrow 2 \quad (26)$$

in which the symbol  $\otimes$  indicates linear convolution,  $\downarrow 2$  decimation by two, and  $r_2$  is the *anti-aliasing* (lowpass) filter.

Define the *enhanced Laplacian pyramid* (ELP)<sup>4</sup> as

$$L_k = G_k - \text{expand}_2\{G_{k+1}\} \quad k = 0, 1, \dots, K-1 \quad (27)$$

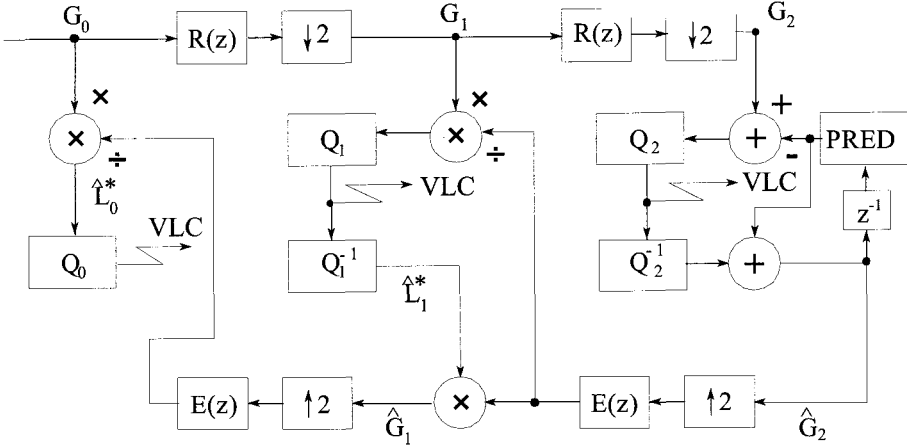


Fig. 4. Flowchart of a hybrid encoder with quantization noise feedback loops on levels 2 (causal DPCM) and 1 (noncausal DPCM). VLC stands for variable length coding.

with

$$\text{expand}_2\{G_{k+1}\} \triangleq (G_{k+1} \uparrow 2) \otimes e_2 \quad (28)$$

in which  $\uparrow 2$  stands for *upsampling* by two, i.e., interleaving samples with zeroes, and  $e_2$  is the interpolation (lowpass) filter, which can be the same as  $r_2$ , apart from a *dc* gain equal to two.

A spatial DPCM can also be *noncausal*, i.e., *interpolation-based*, or *hierarchical*: a coarse image version, i.e., the base band of the GP,  $G_K$ , is encoded followed by the ELP,  $L_k$ ,  $k = K - 1, \dots, 0$ . Quantization error feedback at each layer allows  $L_\infty$  error control via the quantization step at the finest resolution layer<sup>4</sup>.

A *rational* (E)LP (RLP), matching the multiplicative nature of the speckle noise, was defined from the GP (25) and utilized for denoising<sup>7</sup>. The *ratio*, instead of the difference, between the  $k$ th level of the GP and the expanded version of the  $(k+1)$ st level, yields a pyramid

$$L_k^*(i, j) \triangleq \frac{G_k(i, j)}{\tilde{G}_{k+1}(i, j)} \quad (29)$$

in which  $\tilde{G}_{k+1}$  is a shortcoming for  $\text{expand}_2\{G_{k+1}\}$  and the domain of subscripts is the same as for (27). Fig. 3 shows GP and RLP of a test SAR image.

The RLP approximates a *bandpass* image representation, thus retaining all the benefits of multiresolution analysis, including those for data compression. The idea is to causally DPCM encode the small baseband bitmap icon and to quantize and encode the RLP layers<sup>11</sup>. The quantizer step sizes except on the bottom layer may be chosen arbitrarily because of the quantization noise feedback loop at the encoder, which consists of interpolating the same *noisy* data, i.e., affected by the distortion introduced during reconstruction, which will be available at the decoder.

Fig. 4 shows that  $\hat{G}_k$ , the GP that will be reconstructed at the decoder for  $k < K$ , is recursively given by the product of the expanded  $\hat{G}_{k+1}$  and of an approximate version of  $L_k^*$ , namely  $\hat{L}_k^*$ , due to quantization errors, in which  $\hat{L}_k^* = G_k/\hat{G}_{k+1}$ ;  $\hat{L}_K^* = \hat{G}_K$  for the pyramid top or *baseband*.

For the baseband and intermediate layers, linear quantizers are utilized. The step sizes of the linear quantizers are calculated for  $k = K, K - 1, \dots, 1$ , so as to minimize the bit rate for a given distortion, by exploiting the mechanism of quantization noise feedback. Following the procedure reported for the ELP<sup>4</sup>, the entropy-minimizing step sizes are found out to be:

$$\hat{\Delta}_k = \frac{2 \cdot \bar{\sigma}_{k-1}^*}{P_E} \quad k = K, K - 1, \dots, 1 \quad (30)$$

in which  $\bar{\sigma}_{k-1}^*$  is the average standard deviation of  $L_{k-1}^*$  and  $P_E$  is the power gain of the 1-D interpolation filter  $e_2$ .

The last step size,  $\Delta_0$ , as well as the type of quantizer, is crucial when an error bounded encoder is required, since it rules the peak error: *absolute* for the ELP and *relative* for the RLP. In fact,  $G_0$  cannot be exactly recovered from  $G_K$  and  $\hat{L}_k^*$ ,  $k = K - 1, \dots, 1, 0$ , unless quantization on the last layer is extremely fine, which implies a large code rate. Thus, unlike the ELP<sup>4</sup>, the RLP is unsuitable for a strictly lossless compression. Furthermore, the *rational* nature of RLP makes linear quantization unable to guarantee relative-error bounded encoding.

Since the term  $\hat{G}_0$  recursively accounts for previous quantization errors and  $L_0^* = G_0/\hat{G}_0$ , a logarithmic quantization (20) of  $L_0^*$  with a step size  $\Delta_0 \neq 1$  implies that the pixel ratio of original to decoded image is strictly bounded through the step size  $\Delta_0$  of the last quantizer, as in (22). Hence, a relationship identical to (24) is found between the step size of the last quantizer,  $\Delta_0$ , and the dynamic range of relative errors.

## 7. Experimental Results

### 7.1. Multispectral Data

The optical data set comprises a Landsat Thematic Mapper (TM) image, with 8 *bit/pel* and 6 bands out of the 7 available. In fact, the 6th band (thermal infrared) was omitted mainly because of its poor resolution ( $120m \times 120m$  instead of  $30m \times 30m$ ) and scarce spectral correlation with the other bands. The test site is shown in Fig. 5, which portrays a part of the valley of the Adige river, near Trento, in Northern Italy.

TM bands do not span the visible/infrared wavelength interval continuously. Apart from the visible spectrum, the infrared region is coarsely sampled. Thus, all the infrared bands are little correlated, both with the visible bands and with one another. To achieve an optimal multispectral decorrelation, the different bands available should be arranged in a sequence that maximizes the average cross correlation between any couple of consecutive bands<sup>56</sup>.

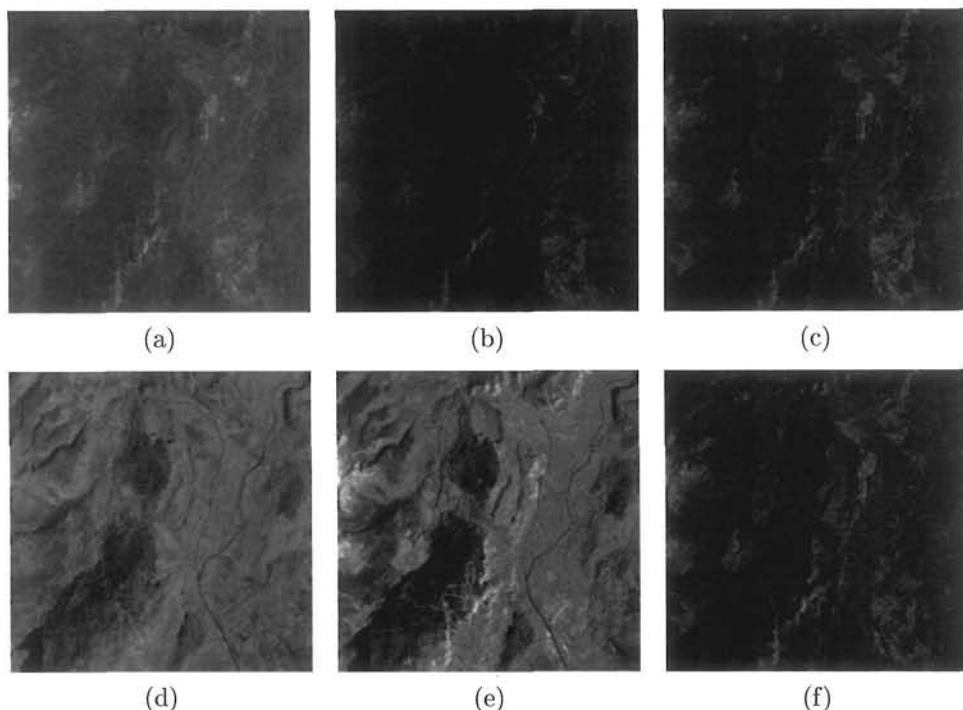


Fig. 5.  $512 \times 512$  details from test TM image: (a) Band 1 (blue), (b) Band 2 (green), (c) Band 3 (red), (d) Band 4 (near infrared), (e) Band 5 (short-wave infrared), (f) Band 7 (middle infrared).

Table 1. Average variance ( $\bar{\sigma}_g^2$ ), estimated noise variance ( $\hat{\sigma}_n^2$ ), SNR ( $dB$ ), and lossless bit rates (in  $bit/pel$ ) of the six 30  $m$  bands of the test TM image achieved by RLPE.

Band (mode)	$\bar{\sigma}_g^2$	$\hat{\sigma}_n^2$	SNR ( $dB$ )	Bit Rate
TM-1 (I)	130.49	1.77	18.67	3.46
TM-2 (B)	52.82	0.37	21.55	1.78
TM-3 (P)	106.72	0.49	23.38	2.61
TM-4 (P)	448.09	5.20	19.35	4.41
TM-5 (B)	769.79	5.57	21.41	3.61
TM-7 (P)	178.44	1.99	19.53	3.27
Avg.	281.06	2.56	20.41	3.19

The optimum *causal* sequence was found to be  $1 \rightarrow 2 \rightarrow 3 \rightarrow 7 \rightarrow 5 \rightarrow 4$ . Band 6, when utilized, is always encoded standalone. A bidirectional, i.e., spectrally noncausal, prediction yields bit rates that are slightly lower, on an average<sup>10</sup>. The optimum bidirectional sequence, was found to be:  $1 \rightarrow 3, 1 \rightarrow 2 \leftarrow 3 \rightarrow 7 \rightarrow 4 \rightarrow 5 \leftarrow 7$ . The difference in rate between causal and noncausal prediction, however, is moderate: the latter provides an average gain of four hundredths of bit per pixel for the optical bands and of nearly eight hundredths for the infrared channels.

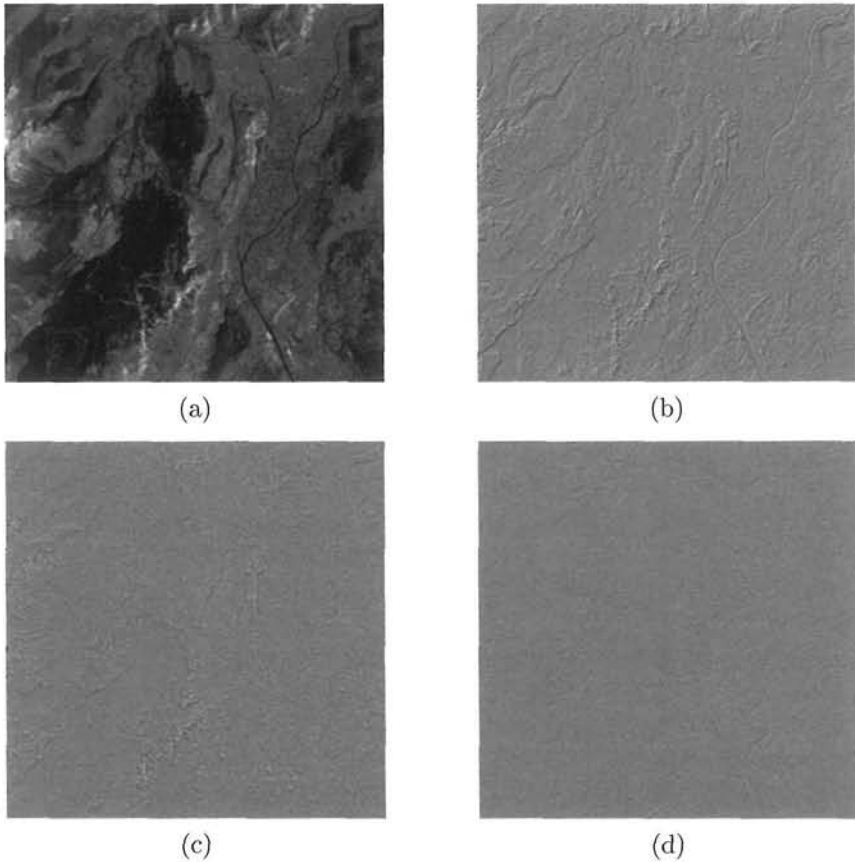


Fig. 6. (a)  $512 \times 512$  test image from Landsat TM band # 5; prediction errors (stretched by 2 and biased by 128 for displaying convenience) produced by: (b) intraband lossless JPEG (predictor # 7); (c) intraband predictor (2-D RLPE); (d) interband predictor (3-D RLPE) with bidirectional prediction from TM bands # 4 and # 7.

Table 7.1 reports the estimated parameters for the six bands, the first of which is encoded in *intra* mode, i.e., without reference to any other previously encoded band. The noise variance is larger in the visible than in the infrared wavelengths; also the signal variance follows such a trend; thus, the intrinsic SNRs are all comparable among the bands.

Bands # 2 and # 5 are bidirectionally predicted to a larger extent than the others bands of the visible and infrared group, respectively. The extremely fitting prediction is demonstrated by the associated prediction residues of Fig. 6(d), which practically comprise the background noise only in most of the image. The work parameters of the algorithm are noncrucial<sup>22</sup> and have been chosen so as to balance coding performances with encoding time (decoding is always real-time).

Rate Distortion plots are reported in Fig. 7(a) for the RLPE scheme, both in *intra*band (2-D RLPE) and in *inter*band mode (3-D RLPE), and for the DCT-based

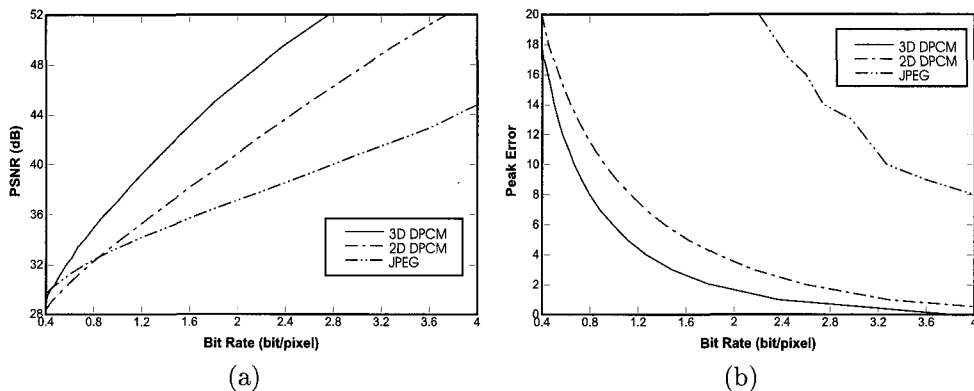


Fig. 7. Band # 5 of TM image (see Fig. 6(a)) compressed by means of RLPE, in *intra*band mode (2-D DPCM) and in bidirectional *inter*band mode (3-D DPCM) from bands # 4 and # 7, and JPEG (*intra*band): (a) PSNR vs. bit rate; (b) peak error vs. bit rate.

lossy JPEG. The test image is the  $512 \times 512$  detail from band # 5 (Fig. 6(a)). In the *interband* mode (3-D) the image is bidirectionally predicted from bands # 4 and # 7. 2-D RLPE gains over JPEG for rates above  $0.8 \text{ bit/pel}$ ; 3-D RLPE crosses the RD plot of JPEG at  $0.4 \text{ bit/pel}$ . The knee for low rates is typical of all causal DPCM schemes and is an effect of quantization noise feedback in the prediction loop. From Fig. 1(a) it appears that, since the “noisy” data reconstructed at the decoder are utilized for prediction, prediction becomes poorer and poorer as the bit rate, and hence the quality of dequantized samples, decreases. The near-lossless performance, shown in the peak error vs. bit rate plots of Fig. 7(b), demonstrates that the two  $L_\infty$ -bounded encoders are far superior to JPEG, which is  $L_2$ -bounded through the user-definable quality factor<sup>46</sup>. The standard deviation of the sensor’s noise was found<sup>19</sup> to be approximately equal to 3 (2.36); hence the *virtually lossless* case, corresponding to a quantization step size  $\Delta = 3$ , with an induced MAD equal to 1, is achieved at  $3.45 \text{ bit/pel}$  by 2-D RLPE and at  $2.43 \text{ bit/pel}$  by 3-D RLPE, with a PSNR gain over JPEG of  $7.5 \text{ dB}$  and  $11 \text{ dB}$ , respectively.

The meaning of the *virtually lossless* term is highlighted in Fig. 8, reporting the bitmaps and the histograms of pixel differences between original and decoded images, for 3-D RLPE and JPEG, respectively, both at  $2.43 \text{ bit/pel}$ . Besides the  $11 \text{ dB}$  PSNR gain of RLPE over JPEG, the error is practically uniformly distributed in  $[-1, 1]$ , as well as uniformly spread over the whole image. In the latter case the error is roughly Gaussian in  $[-15, 17]$  and spatially heterogeneous, much larger around edges and in textured areas than on the background.

## 7.2. Hyperspectral Data

The data set includes also a sequence of hyperspectral images collected in 1997 by the *Airborne Visible InfraRed Imaging Spectrometer* (AVIRIS), operated by NASA/JPL, on the *Cuprite Mine* test site, in Nevada. The sequence is consti-



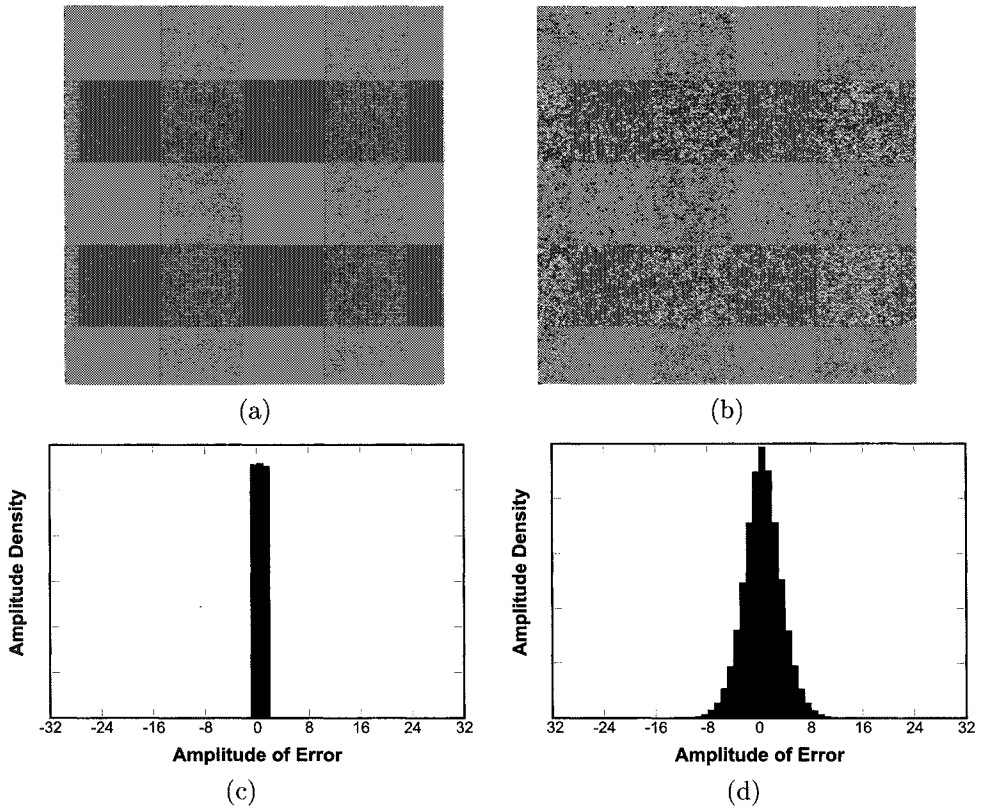


Fig. 8. Error between original and decoded TM band # 5, at same bit rate of 2.43 *bit/pel* (stretched by a factor 7 and biased by 128 for displaying convenience): (a) 3-D RLPE; (b) JPEG; error distribution: (c) 3-D RLPE; (d) JPEG.

tuted by 224 bands recorded at different wavelengths in the range  $380 \div 2500 \text{ nm}$ , with an average spectral separation between two bands of  $10 \text{ nm}$ . The image size is  $614 \times 2048$  pixels. A  $614 \times 512$  subimage was used in this experiment. The raw sequence was acquired by the 12 *bit* analog-to-digital converter (ADC) with which the sensor was equipped in 1995, in place of the former 10-bit ADC. Band # 48 ( $808 \text{ nm}$ ) is shown in Fig. 9. The raw data from the digital counter have been radiometrically calibrated by multiplying by a gain and adding an offset (both varying with wavelengths), and are expressed as radiance values, rounded to integers, and packed in a 16-bit wordlength, including a sign bit. Band # 48 ( $808 \text{ nm}$ ) is shown in Fig. 9. The second spectrometer, covering the near-infrared (NIR) spectrum, was analyzed in a recent work by the authors<sup>14</sup>. It was found that the noise affecting AVIRIS data is somewhat correlated spectrally and across track, and less along track, due to the “wisk-broom” scan mechanism, as well as to postprocessing. Firstly, the 3-D RLPE was run on the test AVIRIS sequence in the reversible mode, i.e., with  $\Delta = 1$ . Each of the 224 bands was decorrelated and encoded with refer-

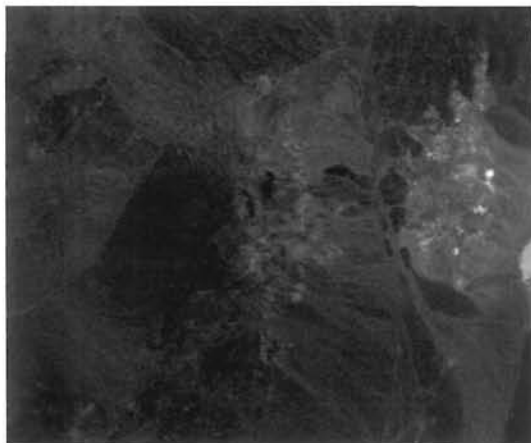


Fig. 9. Band # 48 (808 nm wavelength) of NASA/JPL AVIRIS *Cuprite Mine*, collected in 1997 with size  $614 \times 512$  pixels and a wordlength of 16 bits.

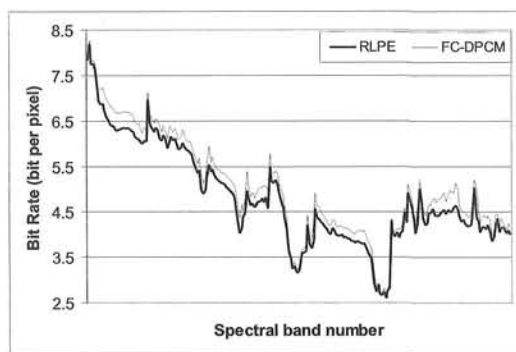


Fig. 10. Bit rates (in *bit/pel*) produced by 3-D RLPE and by the 3-D *fuzzy-clustered* DPCM (FC-DPCM), for the *reversible* encoding of the 224 bands from the test AVIRIS sequences, varying with the wavelength. Each band is predicted both spatially and spectrally from the two previous bands.

ence to its two previous bands. A larger number of bands for prediction is useless<sup>8</sup>, besides being computationally more onerous. Fig. 10 reports the bit rate produced by the encoder varying with the wavelength, for the proposed scheme and for the interband fuzzy-clustered DPCM encoder (3-D FC-DPCM)<sup>8</sup>. The number of predictors  $M = 4$  and the size of prediction support  $S = 18$  are the same for both the encoders. As it appears, the former is slightly superior to the latter, which requires encoding time more than ten times greater, and does not allow real-time decoding.

A distortion analysis varying with coding bit rate was carried out also on AVIRIS data. The test image is band # 48 (808 nm) of *Cuprite '97* and is portrayed in Fig. 9. It is somewhat detailed and richly textured. In fact, on a vegetated area the near-infrared spectrometer captures perhaps the largest amount of spectral information.

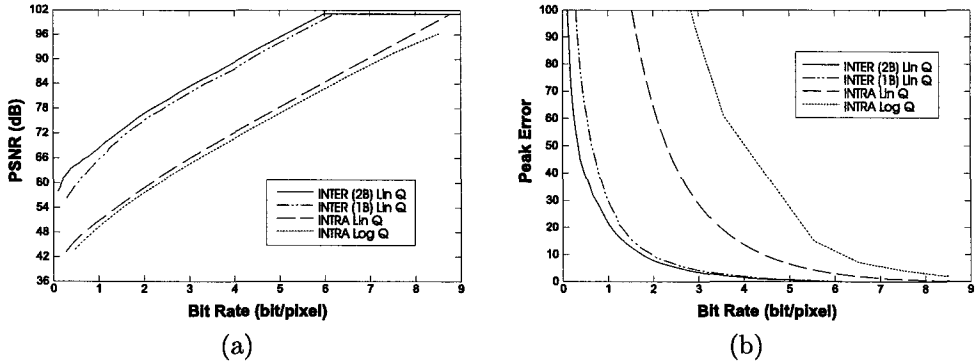


Fig. 11. Band # 48 of AVIRIS *Cuprite '97* compressed by means of the RLPE causal DPCM, in 2-D, i.e., *intra*band mode (INTRA), and 3-D, i.e., *inter*band (INTER) from one (1B) and two (2B) previous bands, with either linear (Lin-Q), or logarithmic quantizer (Log-Q): (a) PSNR vs. bit rate; (b) peak error vs. bit rate.

The bit rates obtained are representative of the average bit rates of the sequence. RD plots are reported in Fig. 11(a) for RLPE operating with  $M = 4$  predictors, in both *intra*band (2-D RLPE) mode ( $S = 8$  coefficients per predictor), and *inter*band mode (3-D RLPE), with reference to either one or two previous bands, with  $S = 12$  and  $S = 16$ , respectively. Due to sign bit, the full scale  $g_{fs}$  in (5) was set equal to  $2^{15} - 1 = 32767$  instead of 65535 (negative values introduced by calibration never occur in the sample band). Hence, the PSNR attains a value of  $10 \log_{10}(12 \cdot g_{fs}^2) \approx 102$  dB, due to integer roundoff noise only, when the reversibility is reached. 3-D RLPE gains 16 through 18 dB over 2-D RLPE, corresponding to almost 3 code bit, depending on whether one or two previous bands are exploited for the 3-D prediction. Two-band 3-D prediction, instead of one-band, gains about 2 dB for medium-high bit rates, and up to 4 dB for low rates. Notice that according to RD theory<sup>38</sup>, when a uniform quantizer is employed, all the SNR/PSNR-bit rate plots are straight lines with slope  $\approx 6$  dB/bit, for rates larger than, say, 1 bit/pel. This does not happen for Log-Q which loses about 2 dB and drifts from the theoretical line as the lossless case is approached. The near-lossless performance is shown in the peak error vs. bit rate plots of Fig. 11(b). Values of peak error are far larger than those reported in Fig. 7(b), because the full scale is now 32767 instead of 255. The trends are in accordance with those of PSNR, except for the Log-Q which achieves a performance much poorer than that of Lin-Q for the *intra* experiment. The standard deviation of the noise was found<sup>14</sup> to be approximately 10; hence, the virtually-lossless case is given by the 3-D encoder (2B) at a bit rate around 3 bit/pel, yielding a compression ratio  $CR > 5$ .

Another experiment concerns assessments of PMAD-constrained coding performances. Bands 35 to 97, covering the NIR wavelengths have been compressed in both MAD-constrained mode (linear quantization) and PMAD constrained mode (logarithmic quantization). The work parameters of RLPE have still been chosen so

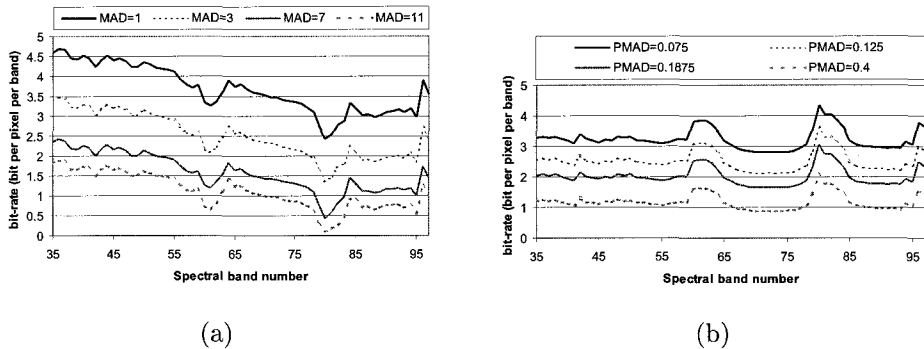
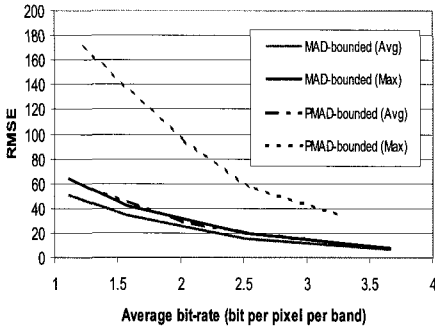


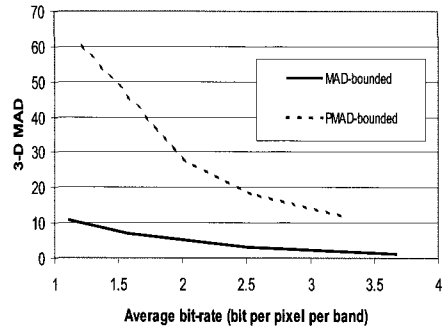
Fig. 12. Bit rates produced by 3-D RLPE on the data produced by the second spectrometer (NIR) of AVIRIS *Cuprite Mine '97*: (a) linear quantization to yield user-defined MAD values; (b) logarithmic quantization to yield user-defined PMAD values.

as to balance performances with encoding time. The outcome bit rates varying with band number, together with the related distortion parameters are shown in Fig. 12. As it appears the bit rate plots follow similar trends varying with the amount of distortion, but quite different trends for the two types of distortion (i.e., either MAD or PMAD). For example, around the water vapor absorption wavelengths ( $\approx$  Band 80) the MAD-bounded plots exhibit pronounced valleys, that can be explained because the intrinsic SNR of the data becomes lower; thus the linear quantizer dramatically abates the *noisy* prediction errors. On the other hand the PMAD-bounded encoder tends to quantize the noisy residues more finely when the signal is lower. Therefore bit rate peaks are generated instead of valleys. More generally speaking, bit rate peaks from the PMAD-bounded encoder are associated with low responses from the spectrometer. This explains why the bit rate plots of Fig. 12(b) never fall below one bit per pixel per band.

Eventually, some of the distortion measures defined in Sect. 2 have been calculated on the distorted hyperspectral pixel vectors achieved by decompressing the bit streams generated by the near-lossless encoder, both MAD- and PMAD-bounded. RMSEs of the vector data, both *average* RMSE (8) and *peak*, i.e., maximum, RMSE (9) are plotted in Fig. 13(a) as a function of the bit rate from the encoder. The MAD-bounded encoder obviously minimizes both the radiometric distortions: average (A-RMSE) and maximum (P-RMSE) Euclidean norm of the pixel error vector. A further advantage is that A-RMSE and P-RMSE are very close to each other for all bit rates. The PMAD-bounded encoder is somewhat poorer: A-RMSE is comparable with that of the former, but P-RMSE is far larger, due to the high-signal components that are coarsely quantized in order to minimize PMAD. Trivially, the MAD of the data cube (12) is exactly equal to the desired value (see Fig. 12(a)), whereas the PMAD, being unconstrained, is higher. Symmetric results, not reported here, have been found by measuring PMAD on MAD-bounded and PMAD-bounded decoded data.

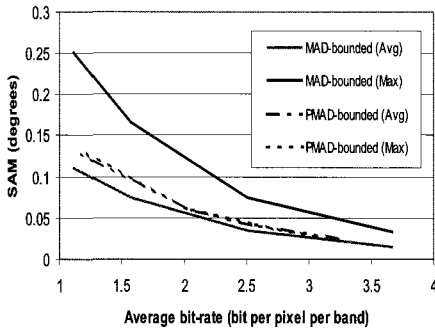


(a)

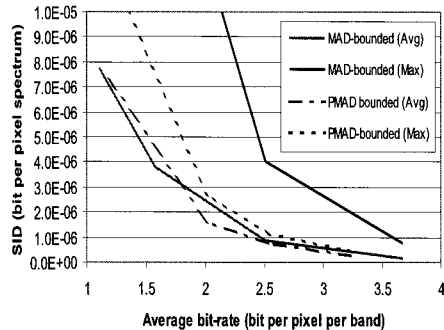


(b)

Fig. 13. Radiometric distortions versus bit rate for compressed AVIRIS *Cuprite Mine '97* data: (a) RMSE; (b) MAD.



(a)



(b)

Fig. 14. Spectral distortions versus bit rate for compressed AVIRIS *Cuprite Mine '97* data: (a) spectral angle mapper (SAM); (b) spectral information divergence (SID).

As far as *radiometric* distortion is concerned, results are not surprising. Radiometric distortions measured on vectors are straightforwardly derived from those measured on scalar pixel values. The introduction of such *spectral* measurements as SAM (14) and SID (18) may overcome the rationale of *distortion*, as established in the signal/image processing community. Fig. 14 shows spectral distortions between original and decompressed hyperspectral pixel vectors. The PMAD-bounded algorithm yields plots (maximum and average SAM in Fig. 14(a)) that lie in the middle between the corresponding ones produced by the MAD-bounded algorithm and are very close to each other too. Since the *maximum* SAM is a better clue of spectral quality of the decoded data than the *average* SAM may be, a likely conclusion would be that PMAD-bounded compression optimizes the *spectral* quality

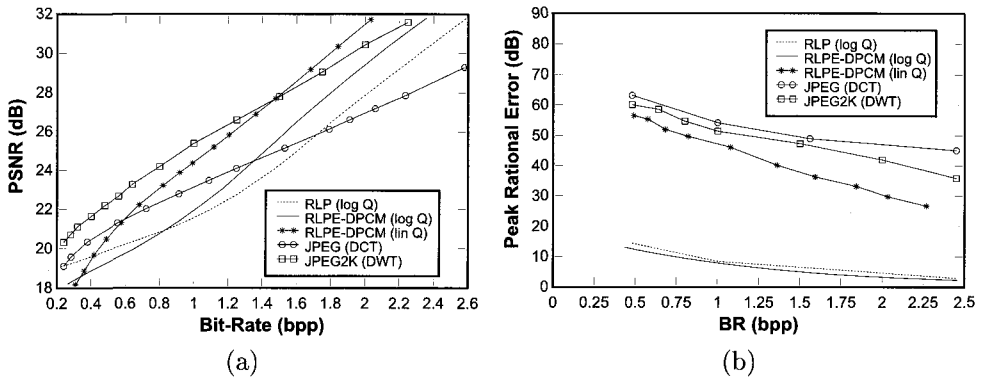


Fig. 15. Comparisons on test SAR image (8 bit/pel): (a) PSNR vs. bit rate; (b) PRE vs. bit rate.

of the data, while MAD-bounded is superior in terms of *radiometric* quality. The considerations expressed for SAM are emphasized by the plots of Fig. 14(b) reporting average and maximum SID. The latter is capable of discriminating spectral quality more finely than SAM does, as previously noticed by Chan<sup>27</sup> in the case of multiband classification.

### 7.3. SAR Data

A thorough performance comparison aimed at highlighting the *rational* near-lossless approach was carried out among causal RLPE DPCM (2-D) with either linear or logarithmic quantization and noncausal DPCM achieved by RLP with linear quantization except on bottom layer where it is logarithmic. The current standard JPEG (block DCT-based)<sup>46</sup> and the upcoming standard JPEG 2000 (wavelet-based)<sup>37</sup> where also considered.

The well known test NASA/JPL AIRSAR image of San Francisco (4 looks amplitude) remapped to 8 bit/pel, was used for the coding experiment. The test SAR image is shown as the bottom layer of the GP in Fig. 3(a).

Fig. 15(a) shows the PSNR vs. the bit rates produced by the five different encoders starting from the test SAR image. For the RLP and the RLPE-DPCM (Log-Q) the rightmost part of the plots (say, bit rate  $> 2$  bit/pel) correspond to a *virtually lossless* coding. As it appears, RLPE DPCM gains more than 1.5 dB PSNR over RLP. The two plots are parallel for medium-high rates and cross each other at approximately 1 bit/pel. This is not surprising because both encoders utilize Log-Q, and it is widely known that noncausal DPCM is preferable to causal DPCM for low rates only<sup>4</sup>. Concerning the schemes utilizing linear quantizers, RLPE-DPCM and JPEG2K share the best Rate Distortion (RD) performances: the former outperforms the latter for rates higher than 1.5 dB, and vice-versa. However, it is evident that the logarithmic quantizer, introduced to allow relative error-bounded coding, yields poorer RD performances than a linear quantizer does, especially for low rates. JPEG and JPEG2K having similar quantizers (psychovisual) follow similar RD trends.

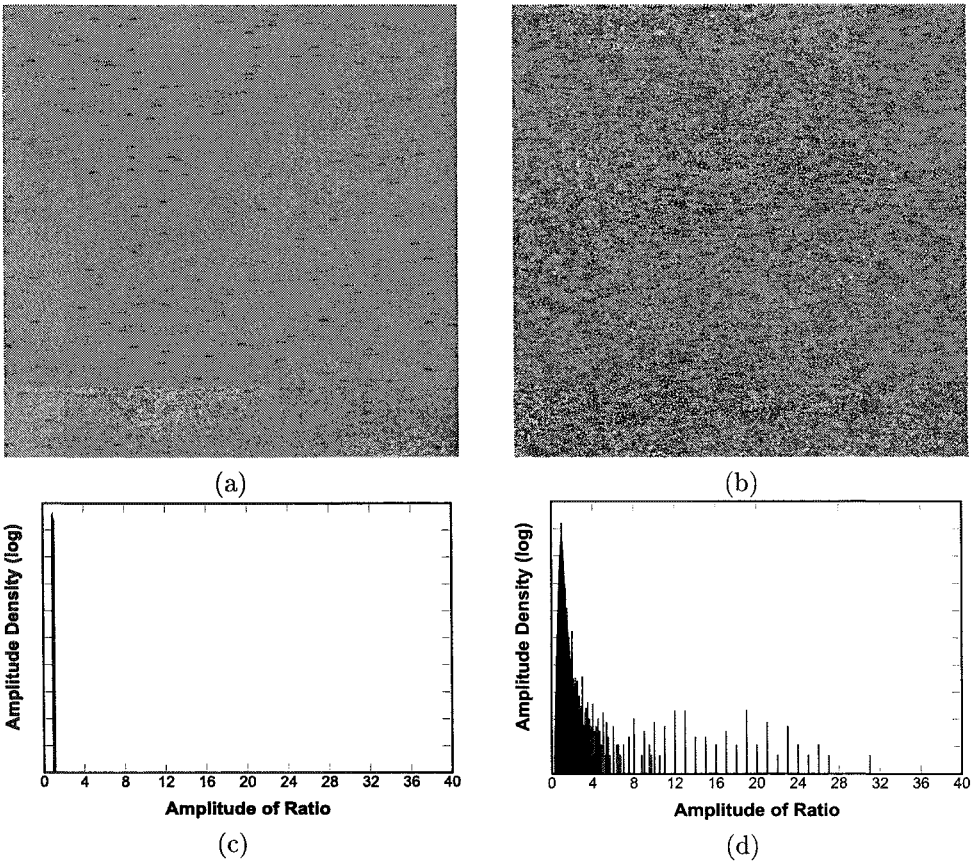


Fig. 16. Ratio of original to reconstructed image at the same bit rate of 2.45 *bit/pel*: (a) RLP; (b) JPEG; distribution of original-to-decoded ratio: (c) RLP; (d) JPEG.

However, the two plots are closer for lower rates and farther apart as the rate increases, unlike what usually happens for “optical”, i.e., noncoherent, images. The visual quality of the decoded images, however, is quite different. For lower rates JPEG2K takes advantages from despeckling the image; instead JPEG introduces severe blocking impairments, especially annoying on the sea.

The relative-error bounded *near-lossless* performance of the two encoder utilizing logarithmic quantization is highlighted in the plots of Fig. 15(b) reporting PRE (23) vs. bit rate between original and decoded versions of the test SAR image. PRE is user defined and far lower than that of the other schemes, including RLPE-DPCM which exhibits the best results among the three encoders using linear quantization. Again RLP and RLPE-DPCM cross each other around 1 *bit/pel*.

The meaning of the *virtually lossless* term is demonstrated in Fig. 16, reporting the bitmaps and the histograms of pixel ratio between original and decoded images, for RLP and JPEG, both at 2.45 *bit/pel*. Although the PSNR gain of RLP over

JPEG is only 2 dB at 2.45 bit/pel, in the former case the relative error is small and uniformly spread; in the latter case it is heterogeneous, much larger around image edges. The variance of the ratio is less than one tenth of that of speckle (nominally 4-look amplitude), for RLP. Hence, the definition of *virtually lossless* applies<sup>12</sup>.

## 8. Conclusions

This work has demonstrated the potential usefulness of *near-lossless* compression, i.e., with bounded pixel error, either *absolute* or *relative*, when it is applied to remote-sensing data. Unlike lossless compression achieving typical CRs around *two*, near-lossless compression can be adjusted to allow a *virtually lossless* compression with CRs larger than *three* for 8-bit multispectral data and larger than *five* for 16-bit hyperspectral data. The main result of this analysis is that, for a given CR, near-lossless methods, either MAD- or PMAD-constrained, are more suitable for preserving the spectral discrimination capability among pixel vectors, which is the principal outcome of spectral information. Therefore, whenever a lossless compression is not practicable, near-lossless compression is recommended in such application where spectral quality is a crucial point. Furthermore, since the maximum reconstruction error is defined by the user before compression, whenever higher CRs are required, the loss of performance expected in application tasks can be accurately modeled and predicted. Eventually, an original approach to *near-lossless* compression of detected SAR images is based on encoding the Rational Laplacian Pyramid of a speckled SAR image, after linearly quantizing its upper layers and logarithmically quantizing its bottom layer, to achieve near-lossless compression with constrained relative error. Besides virtually lossless compression, for which causal DPCM is recommended, noncausal pyramid-based DPCM outperforms causal DPCM when near-lossless compression at low-medium rates is desired.

## References

1. G. P. Abousleman, M. W. Marcellin, and B. R. Hunt. Hyperspectral image compression using entropy-constrained predictive trellis coded quantization. *IEEE Trans. Image Processing*, 6 (4): 566–573, Apr. 1997.
2. A. Abrardo, L. Alparone, and F. Bartolini. Encoding-interleaved hierarchical interpolation for lossless image compression. *Signal Processing*, 56 (2): 321–328, Feb. 1997.
3. B. Aiazzi, L. Alparone, and S. Baronti. A reduced Laplacian pyramid for lossless and progressive image communication. *IEEE Trans. Commun.*, 44 (1): 18–22, Jan. 1996.
4. B. Aiazzi, L. Alparone, S. Baronti, and F. Lotti. Lossless image compression by quantization feedback in a content-driven enhanced Laplacian pyramid. *IEEE Trans. Image Processing*, 6 (6): 831–843, June 1997.
5. B. Aiazzi, L. Alparone, S. Baronti, G. Chirò, F. Lotti, and M. Moroni. A pyramid-based error-bounded encoder: An evaluation on X-ray chest images. *Signal Processing*, 59 (2): 173–187, June 1997.
6. B. Aiazzi, P. S. Alba, L. Alparone, and S. Baronti. Reversible compression of multispectral imagery based on an enhanced inter-band JPEG prediction. In *Proc. IEEE Int. Geoscience And Remote Sensing Symposium*, Singapore, pages 1990–1992, 1997.



7. B. Aiazzi, L. Alparone, and S. Baronti. Multiresolution local-statistics speckle filtering based on a ratio Laplacian pyramid. *IEEE Trans. Geosci. Remote Sensing*, 36 (5): 1466–1476, Sep. 1998.
8. B. Aiazzi, P. Alba, L. Alparone, and S. Baronti. Lossless compression of multi/hyperspectral imagery based on a 3-D fuzzy prediction. *IEEE Trans. Geosci. Remote Sensing*, 37 (5): 2287–2294, Sep. 1999.
9. B. Aiazzi, L. Alparone, S. Baronti, and F. Lotti. Virtually lossless compression of medical images through classified prediction and context-based arithmetic coding. In K. Aizawa, R. L. Stevenson, and Y.-Q. Zhang, Eds., *Visual Communications and Image Processing '99*, Proc. SPIE volume 3653, pages 1033–1040, 1999.
10. B. Aiazzi, L. Alparone, and S. Baronti. Advantages of bidirectional spectral prediction for the reversible compression of multispectral data. In *Proc. IEEE Int. Geoscience And Remote Sensing Symposium*, pages 2043–2045, 1999.
11. B. Aiazzi, L. Alparone, and S. Baronti. Joint compression and de-speckling of SAR images by thresholding and encoding the rational Laplacian pyramid. In *Proc. EU-SAR 2000*, Munich, Germany, pages 2657–2659, 2000.
12. B. Aiazzi, L. Alparone, and S. Baronti. Information-preserving storage of Remote Sensing data: virtually lossless compression of optical and SAR images. In *Proc. IEEE Int. Geoscience And Remote Sensing Symposium*, pages 2657–2659, 2000.
13. B. Aiazzi, L. Alparone, and S. Baronti. Near-lossless image compression: a key to high-quality data distribution. In M. S. Schmalz, Ed., *Mathematics of Data/Image Coding, Compression, and Encryption III*, Proc. SPIE volume 4122, pages 76–85, 2000.
14. B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, and I. Pippi. Information-theoretic assessment of sampled hyperspectral imagers. *IEEE Trans. Geosci. Remote Sensing*, 39 (7): 1447–1458, July 2001.
15. B. Aiazzi, L. Alparone, and S. Baronti. Near-lossless compression of 3-D optical data. *IEEE Trans. Geosci. Remote Sensing*, 39 (11): 2547–2557, Nov. 2001.
16. B. Aiazzi, L. Alparone, S. Baronti, and L. Santurri. Near-lossless compression of multi/hyperspectral images based on a fuzzy matching-pursuits interband prediction. In S. B. Serpico, Ed., *Image and Signal Processing for Remote Sensing VII*, Proc. SPIE volume 4541, pages 252–263, 2001.
17. B. Aiazzi, L. Alparone, and S. Baronti. Relative error-constrained compression for Synthetic Aperture Radar data. In *Mathematics of Data/Image Coding, Compression, and Encryption IV, with Applications*, M. S. Schmalz, Ed., Proc. SPIE volume 4122, pages 128–139, 2001.
18. B. Aiazzi, S. Baronti, and L. Alparone. Near-lossless compression of coherent image data. In *Proc. IEEE Int. Conf. on Image Processing*, Thessaloniki, Greece, pages III: 490–493, 2001.
19. B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, and I. Pippi. Estimating noise and information of multispectral imagery. *J. Optical Engin.*, 41 (3): 656–668, Mar. 2002.
20. B. Aiazzi, L. Alparone, and S. Baronti. Context modeling for near-lossless image coding. *IEEE Signal Processing Lett.*, 9 (3): 77–80, Mar. 2002.
21. B. Aiazzi, L. Alparone, and S. Baronti. Fuzzy logic-based matching pursuits for lossless predictive coding of still images. *IEEE Trans. Fuzzy Systems*, 10 (4): 473–483, Aug. 2002.
22. B. Aiazzi, L. Alparone, and S. Baronti. Near-lossless image compression by relaxation-labelled prediction. *Signal Processing*, 82 (11): 1619–1631, Nov. 2002.
23. Z. Arnavut. Permutations and prediction for lossless compression of multispectral TM images. *IEEE Trans. Geosci. Remote Sensing*, 36 (3): 999–1003, May 1998.

24. J. A. Benediktsson, J. R. Sveinsson, and K. Arnason. Classification and feature extraction of AVIRIS data. *IEEE Trans. Geosci. Remote Sensing*, 33 (5): 1194–1205, Sep. 1995.
25. A. Benazza-Benyahia, J.-C. Pesquet, and M. Hamdi. Vector lifting schemes for lossless coding and progressive archival of multispectral images. *IEEE Trans. Geosci. Remote Sensing*, 40 (9): 2011–2024, Sep. 2002.
26. B. Carpentieri, M. J. Weinberger, and G. Seroussi. Lossless compression of continuous-tone images. *Proc. of the IEEE*, 88 (11):1797–1809, Nov. 2000.
27. Chein-I. Chang. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Trans. Inform. Theory*, 46 (5): 1927–1932, Aug. 2000.
28. K. Chen and T. V. Ramabadran. Near-lossless compression of medical images through entropy-coded DPCM. *IEEE Trans. Medical Imaging*, 13 (3): 538–548, 1994.
29. *Lossless Data Compression: Recommendation for Space Data Systems Standards (Blue Book)*. Consultative Committee for Space Data Systems, Washington, D. C.: CCSDS, May 1997.
30. *Lossless Data Compression: Report Concerning Space Data Systems Standards (Green Book)*. Consultative Committee for Space Data Systems, Washington, D. C.: CCSDS, May 1997.
31. G. Deng, H. Ye, and L. W. Cahill. Adaptive combination of linear predictors for lossless image compression. *IEE Proc.-Sci. Meas. Technol.*, 147 (6): 414–419, Nov. 2000.
32. D. Gleich, P. Planinsic, B. Gergic, and Z. Cucej. Progressive space frequency quantization for SAR data compression. *IEEE Trans. Geosci. Remote Sensing*, 40 (1): 3–10, Jan. 2002.
33. F. Golchin and K. K. Paliwal. Classified adaptive prediction and entropy coding for lossless coding of images. In *Proc. IEEE Int. Conf. on Image Processing*, volume III/III, pages 110–113, 1997.
34. S. W. Golomb. Run-length encoding. *IEEE Trans. Inform. Theory*, 12: 399–401, July 1966.
35. J. W. Goodman. Some fundamental properties of speckle. *J. Opt. Soc. Amer.*, 66 (11): 1145–1150, Nov. 1976.
36. ISO 15887-2000: Space data and information transfer systems – Data systems – Lossless data compression. ISO TC 20/SC 13/ICS 49.140, 12-10-2000.
37. ISO/IEC FCD 15444-1: Information technology–JPEG 2000 image coding system: Core coding system [WG 1 N 1646]. ISO/IEC JTC 1/SC 29/WG1, March 2000.
38. N. S. Jayant and P. Noll. *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Prentice Hall, Englewood Cliffs, NJ, 1984.
39. S. G. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, New York, 1999.
40. I. Matsuda, H. Mori, and S. Itoh. Lossless coding of still images using minimum-rate predictors. In *Proc. IEEE Int. Conf. on Image Processing*, volume I/III, pages 132–135, 2000.
41. N. D. Memon, K. Sayood, and S. S. Magliveras. Lossless compression of multispectral image data. *IEEE Trans. Geosci. Remote Sensing*, 32 (2): 282–287, Mar. 1994.
42. N. Merhav, G. Seroussi, and M. J. Weinberger. Coding of sources with two-sided geometric distributions with unknown parameters. *IEEE Trans. Inform. Theory*, 46 (1): 229–236, 2000.
43. S.-E. Qian, A. B. Hollinger, and Y. Hamiaux. Study of real-time lossless data compression for hyperspectral imagery. In *Proc. IEEE Int. Geoscience And Remote Sensing Symposium*, Hamburg, Germany, pages 2043–2045, 1999.

44. T. V. Ramabadran and K. Chen. The use of contextual information in the reversible compression of medical images. *IEEE Trans. Medical Imaging*, 11 (2): 185–195, Mar. 1992.
45. S. D. Rane and G. Sapiro. Evaluation of JPEG-LS, the new lossless and controlled-lossy still image compression standard, for compression of high-resolution elevation data. *IEEE Trans. Geosci. Remote Sensing*, 39 (10): 2298–2306, Oct. 2001.
46. K. K. Rao and J. J. Hwang. *Techniques and Standards for Image, Video, and Audio Coding*. Prentice Hall, Engl. Cliffs, NJ, 1996.
47. J. Reichel, G. Menegaz, M. J. Nadenau, and M. Kunt. Integer wavelet transform for embedded lossy to lossless image compression. *IEEE Trans. Image Processing*, 10 (3): 383–392, Mar. 2001.
48. R. F. Rice and J. R. Plaunt. Adaptive variable-length coding for efficient compression of spacecraft television data. *IEEE Trans. Commun. Technol.*, COM-19 (6): 889–897, Dec. 1971.
49. J. Rissanen. A universal data compression system. *IEEE Trans. Inform. Theory*, IT-29: 656–664, Sept. 1983.
50. R. E. Roger and J. F. Arnold. Reversible image compression bounded by noise. *IEEE Trans. Geosci. Remote Sensing*, 32 (1): 19–24, Jan. 1994.
51. R. E. Roger and M. C. Cavenor. Lossless compression of AVIRIS images. *IEEE Trans. Image Processing*, 5 (5): 713–719, May 1996.
52. M. J. Ryan and J. F. Arnold. Lossy compression of hyperspectral data using vector quantization. *Remote Sens. Environ.*, 61 (3): 419–436, Sep. 1997.
53. A. Said and W. A. Pearlman. An image multiresolution representation for lossless and lossy compression. *IEEE Trans. Image Processing*, 5 (9): 1303–1310, Sep. 1996.
54. J. T. Tilton and M. Manohar. Preserving radiometric resolution in remotely sensed data with lossy compression. *IEEE Trans. Geosci. Remote Sensing*, 35 (5): 1171–1176, Sep. 1997.
55. V. D. Vaughn and T. S. Wilkinson. System considerations for multispectral image compression design. *IEEE Signal Processing Magazine*, 12 (1): 19–31, Jan. 1995.
56. J. Wang, K. Zhang, and S. Tang. Spectral and spatial decorrelation of Landsat-TM data for lossless compression. *IEEE Trans. Geosci. Remote Sensing*, 33 (5): 1277–1285, Sep. 1995.
57. M. J. Weinberger, J. J. Rissanen, and R. B. Arps. Applications of universal context modeling to lossless compression of gray-scale images. *IEEE Trans. Image Processing*, 5 (4): 575–586, Apr. 1996.
58. M. J. Weinberger, G. Seroussi, and G. Sapiro. The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS. *IEEE Trans. Image Processing*, 9 (8): 1309–1324, Aug. 2000.
59. I. H. Witten, R. M. Neal, and J. G. Cleary. Arithmetic coding for data compression. *Commun. ACM*, 30: 520–540, June 1987.
60. X. Wu and N. Memon. Context-based, adaptive, lossless image coding. *IEEE Trans. Commun.*, 45 (4): 437–444, Apr. 1997.
61. X. Wu and P. Bao.  $L_\infty$  constrained high-fidelity image compression via adaptive context modeling. *IEEE Trans. Image Processing*, 9 (4): 536–542, Apr. 2000.
62. X. Wu and N. Memon. Context-based lossless interband compression—Extending CALIC. *IEEE Trans. Image Processing*, 9 (6): 994–1001, Jun. 2000.
63. Z. Zeng and I. G. Cumming. SAR image data compression using a tree-structure wavelet transform. *IEEE Trans. Geosci. Remote Sensing*, 39 (3): 546–552, Mar. 2001.
64. J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Trans. Inform. Theory*, IT-23 (3): 337–343, May 1977.

## **Emerging Topics**

This page is intentionally left blank

## CHAPTER 24

### IMAGE ENHANCEMENT IN GROUND PENETRATING RADAR SIGNAL PROCESSING TO DETECT LANDMINES

Xiaoyin Xu and Eric L. Miller

*Center for Subsurface Sensing and Imaging Systems  
Northeastern University  
Boston, Massachusetts, USA*

The use of ground penetrating radar (GPR) array for detecting and localizing buried objects has received considerable attention in recent years in areas such as landmine and unexploded ordnance remediation, utility line mapping, and archaeological discovery. A typical GPR array is implemented by moving a transmitter and receiver along a linear track. At every stop of the system, the transmitter emits a short pulse of electromagnetic energy which interacts with the surrounding medium. Based on observations of scattered fields collected by the array the objective of the problem is to determine if an object is present in the field of view of the array and, furthermore, to localize its position.

From the perspective of image enhancement, we can apply image processing methods such as histogram modifications to improve the quality of GPR imagery. Histogram modification is a family of point operation methods that modify the pixels to enhance the contrast of the images. Using histogram modification methods to process GPR images can enhance the landmine reflected signals, which are usually weaker than the specular ground reflection. Enhanced landmine reflected signals allow better detection and localization. On the other hand, histogram modification inevitably generates noise and some undesirable artifacts in the image. To remove the noise and artifacts, we use median filtering and adaptive filtering techniques to remove the noise, usually in the form of speckle noise, and the artifacts, usually in the form of horizontal streaks. The resulting image then has enhanced landmine reflected signals and an approximate homogeneous background, which allows better detection of buried landmines.

#### 1. Introduction

The problem of detecting and localizing buried landmines has received considerably amount of research interest in recent years. Subsurface sensing techniques such as GPR<sup>1,2</sup>, infrared imaging<sup>3,4</sup>, electromagnetics/magnetics<sup>5,6</sup>, laser-induced acoustic imaging<sup>7,8</sup>, and nuclear quadrupole resonance<sup>9</sup> have been applied to landmine detection. Among these techniques, GPR is widely used because it is sensitive to variations of all three electromagnetic parameters of a medium, i.e., electrical conductivity, electrical permittivity, and magnetic permeability. Therefore GPR is able

to detect both metallic and non-metallic objects. In addition GPR is able to localize buried objects in a three-dimensional space and obtain discrimination information based on the received signals. In spite of the advantages of GPR, proper signal processing techniques are needed to extract information from GPR data because there is clutter in the GPR signals. Here clutter is defined as any unwanted signals in the received GPR data. Clutter can come from the reflection by the ground surface and scattering in the soil.

In this section, we give a brief description of GPR and its operation principles. Readers who are interested in detailed GPR description are referred to the book by Daniels<sup>10</sup>. GPR can be defined as a radar whose goal is to detect and identify structures underground<sup>1</sup>. It has been in use for about 20 years in contrast to the more than 50 year history of conventional radar<sup>1</sup>.

A simple diagram of a GPR is given in Fig. 1(a). It consists of a transmitting antenna (transmitter) and a receiving antenna (receiver). The transmitter radiates a pulse into the ground and the receiver collects the echo for a certain time period, usually 10 to 20 ns. The transmitted pulse may be any transient signal, sine wave, steps, and Gaussian waves are all possible. Pulse widths are usually in the order of a few nanoseconds<sup>1</sup>. Compared with conventional radar, GPR has a much broader bandwidth, from a few megahertz to a few gigahertz. GPR operates at high frequency so that it can provide images of high resolution of subsurface reflectivity. At such high frequencies, electromagnetic waves are extremely vulnerable to interference. This interference can be reduced by averaging a number of GPR returns obtained as a function of time without moving the GPR<sup>1</sup>. The most critical part of a GPR is the transmitting and receiving antennae. The antennae are usually in the form of dipoles, which are heavily loaded to reduce as much as possible antenna ringing. Usually the transmitter and receiver are separated, i.e., the GPR is a bistatic radar. For typical GPR, the receiver takes the forms of a sampling system. The sampling system constructs the received signal from discrete samples of successive periods of the reflected waveform. "The sampling system makes it possible to use an amplifier with time- (or range-) dependent gain controlled by a computer prior to the sampler in order to minimize sampling noise"<sup>1</sup>.

Field operation of a GPR system is very simple. Commonly, a vehicle mounted GPR system (VMGPR) surveys an interested area step by step along a linear track. A VMGPR may consist of a single GPR or an array of GPRs. At each step, the VMGPR operates in the following sequence: 1) the transmitter radiates a pulse into the ground and turns off, 2) the receiver turns on to collect reflected signal, 3) the receiver turns off after a short time, usually 10 to 20 ns. The data that are recorded by one receiver at one step is called a trace (or a time-series). Fig. 1(b) shows a typical received GPR signal. Denote the direction of VMGPR movement as  $x$ , then the VMGPR will collect reflections in the  $(x, t)$  plane. If the VMGPR is a GPR array, a three-dimensional image of the subsurface will be generated in the  $(x, y, t)$  space. Depending on the task to perform, the step size of the VMGPR

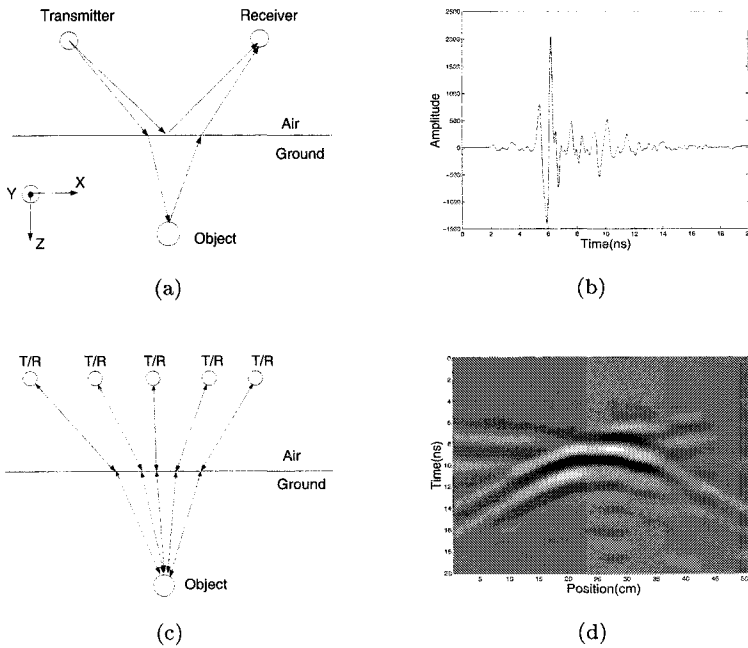


Fig. 1. GPR and its signal, (a) a single GPR, (b) a single trace of received GPR signals, (c) a GPR array, each circle represents a pair of transmitters and receivers, (d) image obtained by the GPR array after subtracting background. Note that for the purpose of illustration we chose to use a GPR image with a strong landmine reflected signals, i.e., the manifest hyperbolic curve seen in the image. The landmine reflected signals are so obvious in this image there is little need of image processing. However, in most cases, GPR image contains a much weaker landmine reflected signals and therefore image processing are needed to enhance the image.

can vary from a few centimeters to a fractions of meters. Fig. 1(c) and (d) show a GPR array and its received signals. At positions close to the object, the array registers stronger reflections at small time-delays. At farther away positions, the reflections are weaker and time-delays are large. Thus typical GPR signals show a manifest hyperbolic curve around the position of an object. Fig. 1(d) confirms that the object (landmine) reflected signals are transient in two directions, the signals are clearly visible between 15 to 35 cm in the horizontal position and 6 to 14 ns in the time axis.

GPR is different from the conventional radar in that 1) GPR operates in near-field scenario, 2) most GPR antenna are close to the air/ground interface and ground reflection cannot be ignored, 3) the medium GPR operates in is highly lossy. At a frequency of  $f$ , the attenuation factor is

$$\alpha = \frac{1}{2}\sigma(f)\sqrt{\frac{\mu}{\epsilon}} \quad (1)$$



where  $\mu$  is the permeability,  $\epsilon$  is the permittivity, and  $\sigma$  is the conductivity of the medium. Note that the conductivity is frequency-dependent. Depending on the shape of antenna and the way it works, a spreading function may also be used to model GPR signal accurately.

Because soil is usually inhomogeneous and air/ground interface is rough, there are multipath reflection and refraction in GPR signal. When one includes frequency dependent conductivity at the soil, it becomes prohibitive to process the GPR signal based solely on physics principles. On the other hand, we can overcome the difficulty in modeling by exploiting the data diversity provided by a GPR array and use some basic physics principles and image processing methods to look for some specific features in the GPR signals.

## 2. Image Processing Methods To Enhance Detection from GPR

### Data

A GPR trace consists of measurement noise, specular reflection from ground surface, clutter, and possibly object reflected signals. Measurement noise comes from imperfections of the GPR hardware, approximation in analog to digital conversion, and human error. Specular reflection is the electromagnetic waves which have bounced off of the air-ground interface. Specular reflection can be reduced by some GPR configurations such as forward-looking GPR<sup>11</sup>, but it cannot be totally eliminated. In GPR signal processing the biggest challenge is to remove the specular reflection. In practice, a GPR moves along a linear track to collect a series of traces in order to find subsurface objects. A synthesized GPR image consists of a number of traces obtained at different positions.

Here we investigate using histogram modification methods to enhance GPR imagery. Specular reflections from the air-ground interface are the strongest components in the GPR images. Compared with the specular reflection, landmine reflect signals are of low amplitude and hard to observe. Histogram modification is very useful in enhancing weak signals<sup>12</sup>. Using iterative contrast stretch and background removal method, we can obtain better GPR images to improve landmine detection. Examples are given in Section 3.2 and 4.2.

The histogram of a signal represents the relative frequency of occurrence of the various amplitude levels in the signal. Obviously the signal can be of one-dimensional or multi-dimensional. In GPR image enhancement, histogram modeling techniques modify a GPR image so that its histogram has a desired shape. For example, in most cases, we would like to enhance the low-contrast part of the image which contains the landmine reflect signals. Histogram modeling techniques fall into three categories, contrast stretch, histogram modification, and histogram specification, respectively<sup>12</sup>. Here we use contrast stretch to obtain better GPR images in order to improve detection of landmines.

At this point, some observation about landmine detection are helpful to understand the challenges of the problem. To the contrary of intuition, it is usually the

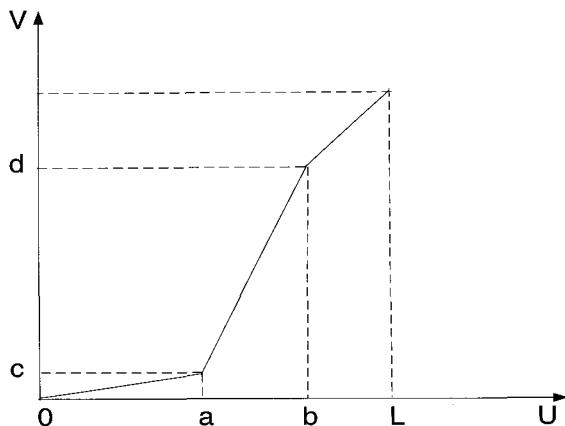


Fig. 2. Contrast stretch enhances the contrast of mid-region pixel values. Pixel values in the mid-region  $[a, b)$  are given a larger contrast range, whereas pixel values outside the mid-region are compressed to a small range.

shallowly buried landmines that are difficult to detect because they are close to the surface and their signals are mixed with the specular reflection from the ground. For the deep buried landmines, time-gating is often sufficient to separate the landmine reflected signal from the ground bounce, i.e., by throwing away signals arrived in the first few nanoseconds one will have only the landmine reflected signal left and detection is not very difficult.

Returning to the detection of shallowly buried landmines, we note that simple background subtraction does not always produce satisfactory results (c.f. Fig. 4, 7, and 8). On the other hand, using contrast stretch, we can enhance the landmine reflected signal and then background subtraction will be more constructive. Better result can be obtained by iteratively applying the contrast stretch and background removal (c.f. Fig. 7 and 8). In the next section, we introduce an iterative contrast stretch method to enhance GPR images. The field data used in this chapter are collected by BRTRC, Inc. at Fort AP Hill test site in Virginia, USA.

Fig. 2 shows how the contrast stretch increases the contrast of mid-region pixel values of an image  $U$ . Mathematically, contrast stretch changes the contrast of the image  $U$  by

$$V = \begin{cases} aU & 0 < U < a \\ b(U - a) + c & a \leq U < b \\ \gamma(U - b) + d & b \leq U \end{cases} \quad (2)$$

where  $a$  and  $b$  are the lower and upper boundaries defining the mid-region. Pixel values falling into  $[a, b)$  will be given a large contrast region, making them more observable. For pixel values less than  $a$  or greater than  $b$ , they are given a smaller range of contrast. The boundaries can be chosen in an ad-hoc manner or by an

iterative method, which we will introduce later on. In specific cases, those pixel values are set to zero and 255 as we will do in the GPR image processing. The reason for setting pixel values smaller than  $a$  to zero and larger than  $b$  to 255 is that we are interested in enhancing the landmine reflected signals only, which fall into the mid-region. In this case, the slope of the mid-region is determined from Fig. 2.

### 3. Iterative Contrast Stretch with Median Filtering

#### 3.1. Method

In landmine detection using GPR, the desired signal is the reflected wave from buried landmines. To enhance the landmine reflected signals, we use contrast stretch to increase the contrast of the GPR image. After the contrast stretch, background removal is carried out by subtracting the ensemble average of the equalized image and a median filter is applied to remove any speckle point in the image. The whole process is iterated to give better result before it diverges. The flowchart of iterative contrast stretch (ICS) is shown in Fig. 3. The median filter is taken over a window of  $3 \times 3$ . For a received GPR image  $\mathbf{y}(m, n)$  of size  $M \times N$ , the mean subtraction is carried out as

$$\mathbf{z}(:, n) = \mathbf{y}(:, n) - \bar{\mathbf{y}}, \quad n = 1, \dots, N \quad (3)$$

where  $\bar{\mathbf{y}} = \frac{1}{N} \sum_{n=1}^N \mathbf{y}(:, n)$  is the ensemble average of  $\mathbf{y}$  along the column direction. The median filter is defined as

$$\mathbf{z}(m, n) = \text{median}\{\mathbf{y}(m - k, n - l), (k, l) \in W\} \quad (4)$$

where  $W$  is a pre-chosen filter window, usually of size  $3 \times 3$ ,  $5 \times 5$ , or  $7 \times 7$ . The median filter is able to remove a single very unrepresentative pixel in the filter window. The algorithm for median filtering requires arranging the pixel values in the filter window in ascending or descending order and picking the middle value. If the number of pixels in the filter is an even-number  $2k$ , the median value is calculated as the average of the arranged pixels at position  $k$  and  $k + 1$ . Because the average of two numbers is usually different from either of the two numbers, median filter is almost always taken over a filter window consisting of an odd-number of pixels such that the median is an actual pixel value in the filter window. For this reason the median filter is better at preserving sharp discontinuities. Because in increasing image contrast, the histogram stretch does not discriminate the desired details from the landmine reflected signals and other small undesirable disturbance, interference will appear as the iteration goes on. The interference appears in the form of speckle noise and can be well removed by the median filter. The whole process can be stopped after a few iterations. In the next section, we use examples to compare the performance of ICS with and without median filtering.

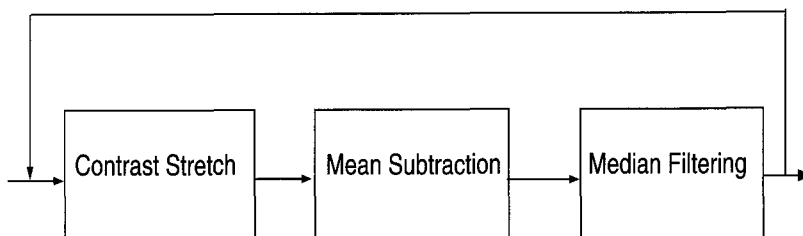


Fig. 3. Flowchart of iterative contrast stretch.

## 3.2. Examples

### 3.2.1. 2-D examples

In this section we use some examples from field data to show the performance of the ICS. Fig. 4(a) displays a GPR image from a buried EM12 landmine at depth of 2.54 cm (1 inch). Output of subtracting the ensemble average is shown in Fig. 4(b). Fig. 4(c) and (d) plot the output of energy detector, i.e., columnwise  $L_2$  norm, of Fig. 4(a) and (b). Energy detector essentially compares the energy of each column to a preset threshold. If the threshold is exceeded, a landmine detection is declared. Otherwise the detector proceeds to the next position to process more data. Comparing Fig. 4(c) and (d) we see that there is not too much improvement. Results of contrast stretch are shown in Fig. 5 where the images generated with and without median filtering are compared. It is seen that in both cases the landmine is readily observable after 6 iterations but the images generated without median filtering have speckle noise. The stripes to the left and right side of the landmine signal are due to the background removal method, which is a subtraction of the ensemble average.

Fig. 6 shows how median filtering can help to improve detection. Fig. 6(a) shows the output of energy detector without median filtering. Fig. 6(b) shows the output of energy detector with median filtering. It is seen that applying median filtering helps to lower the noise floor in Fig. 6(a) and therefore reduce the probability of false-alarm in detection. Fig. 7 shows results of applying the ICS over a buried M21 landmine. It is seen that as iteration increases, the ICS produces better images. Another example is shown in Fig. 8. In the figure, a VS-1.6 landmine is buried at 2.54 cm. The landmine is easily observable in Fig. 8 (d). The bright spot at the upper left corner can be eliminated because in most cases we know the height of the GPR and signals arrived much earlier than the ground bounce can be safely ignored.

### 3.2.2. 3-D examples

The ICS can be expanded to three dimensions in a straight-forward manner. Again we use data collected at Fort AP Hill test site to demonstrate the performance of the ICS. The data were collected by a single GPR running down a linear track, starting

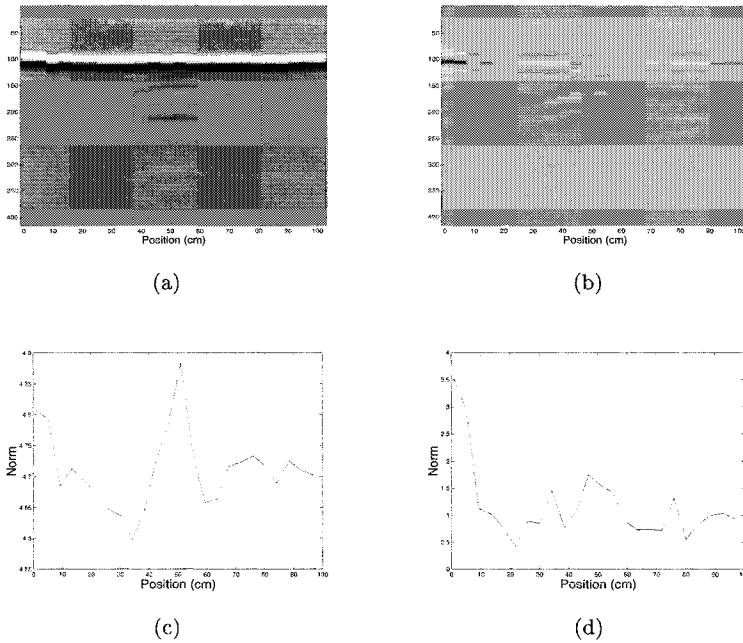


Fig. 4. (a) Original GPR image over an EM12 landmine, (b) image after subtracting the ensemble average, (c) columnwise  $L_2$  norm of (a), (d) columnwise  $L_2$  norm of (b).

from different cross-track position. The whole data set is of size  $416 \times 21 \times 8$  or  $416 \times 24 \times 8$ , which means that at each stop the GPR takes 416 samples in time and makes 21 or 24 stops along the track, covering roughly one meter to 1.2 meters. In both cases, the GPR starts again in one of eight different cross-track positions. Two examples are given in Fig. 9 and 10, over an M15 and a VS-2.2 landmine. While in both cases there are speckle noises in some frames, it is seen that by combining the eight images we can easily detect the landmine.

## 4. Contrast Stretch with Entropy Regularization

### 4.1. Method

The percentage linear contrast stretch uses a specified minimum and maximum values that lie in a certain percentage of pixels from the mean of the histogram. Applying the contrast stretch in GPR image will enhance the contrast of the object reflected signals and make them easy to detection. Consider an image with a minimum brightness value of 0 and a maximum value of 255. Assume we choose a lower boundary of 10 and an upper boundary of 200 in the contrast stretch, the pixels in the original image with a pixel value less than 10 or greater than 200 are changed to 0 and 255, respectively. And the image is linearly stretched between 0

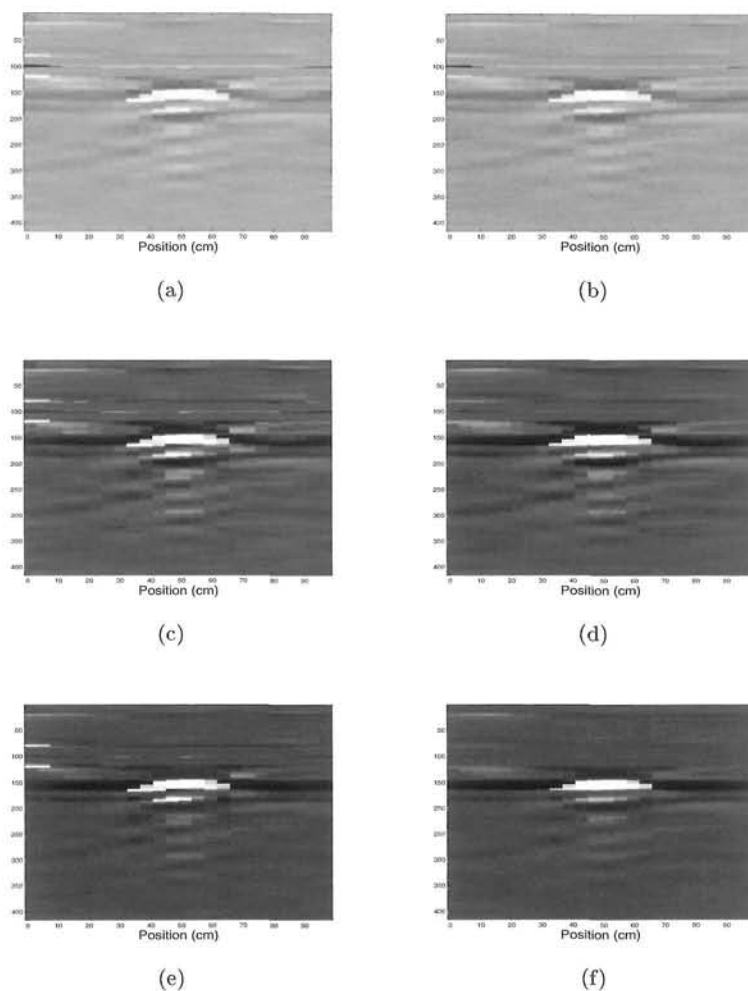


Fig. 5. Left column, images of the ICS without median filtering at iteration 1, 3, and 6. Right column, images of the ICS with median filtering at iteration 1, 3, and 6.

and 255. Therefore, part of the original image that is of low-contrast now is of higher contrast. The difficulty in the percentage linear contrast stretch is to determine the lower and upper percentage boundaries. Boundaries that are too far apart will have little effect on the GPR images while boundaries too close will inadvertently enhance a lot of details of the GPR images while most of them are noise and clutters. Therefore it is important to have a criterion in choosing the proper boundaries. We can write the CS as an operator such that

$$Y = CS(X, a, b) \quad (5)$$

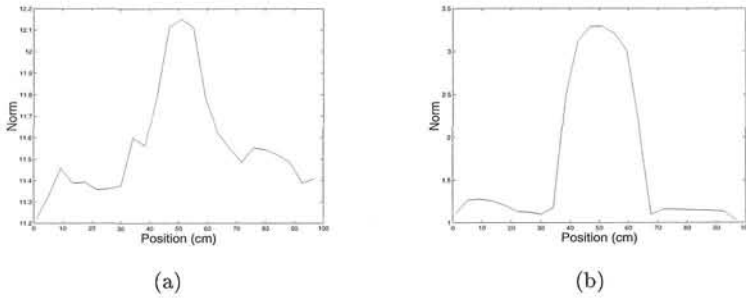


Fig. 6. (a) Norm of the ICS result without median filtering at iteration 6, (b) norm of the ICS result with median filtering at iteration 6. The landmine is buried at position 50 cm.

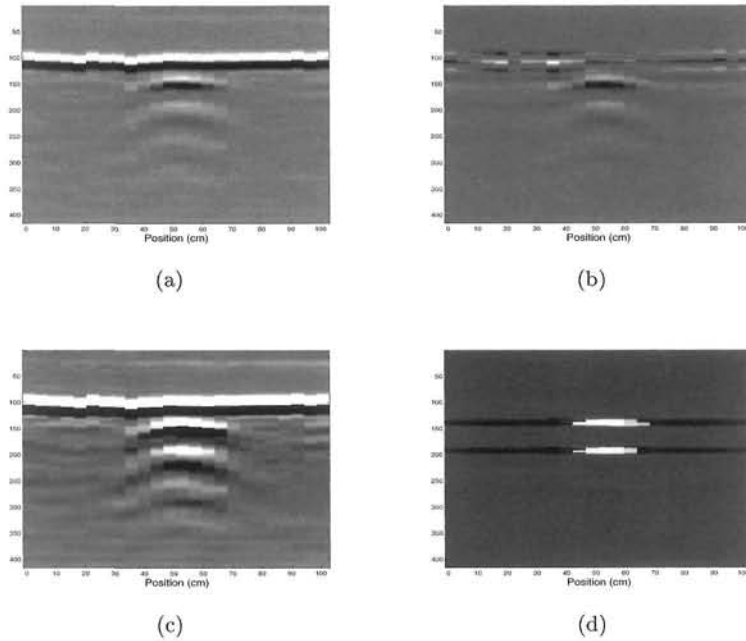


Fig. 7. (a) Original GPR image over an M21 landmine, (b) image after subtracting the ensemble average, (c) image of contrast stretch, (d) image of the ICS result at iteration 6.

where  $Y$  is the resulting image of size  $M \times N$  and  $a$  and  $b$  are the lower and upper boundary, respectively, and  $0 \leq a < b \leq 255$ .

For the purpose of detection, we look for the CS that decomposes a GPR image into three parts, the specular reflection, the object reflected signals and possibly a homogeneous background. A measurement of the homogeneity of an image is its entropy. Entropy measures amount of information carried in an image. A large

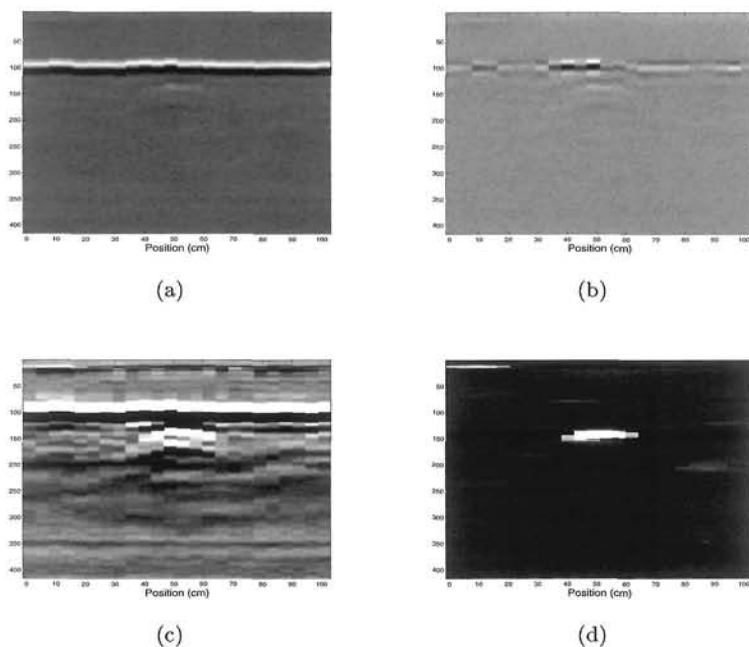


Fig. 8. (a) Original GPR image over a VS-1.6 landmine, (b) image after subtracting the ensemble average, (c) image of contrast stretch, (d) image of the ICS at iteration 6.

entropy corresponds to a large amount of information and vice versa. Intuitively, a homogeneous image has a smaller entropy than an inhomogeneous image. Assume that the image  $X$  of size  $M \times N$  takes discrete values  $1, 2, 3, \dots$  with respective probabilities  $p_1, p_2, p_3, \dots$ . Then the entropy is given by

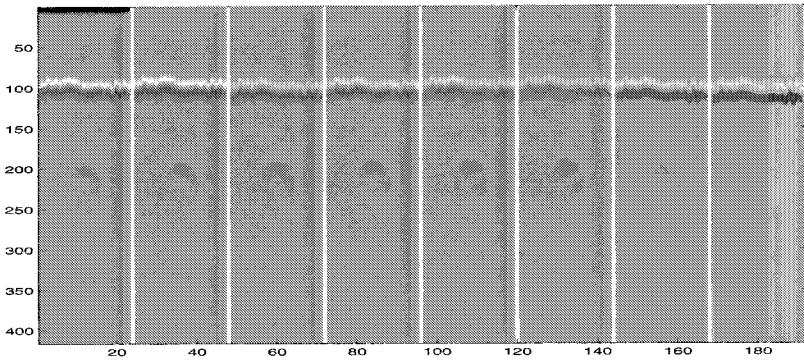
$$E(X) = - \sum_i p_i \log p_i. \quad (6)$$

Calculating the entropy using (6) has two computational problems, first, it is time-consuming to compute the  $p_i$  for a large image, second, if the discrete values of  $X$  are predetermined, then there is possibility that  $p_i = 0$  for some  $i$  and logarithm of zero cannot be computed. To circumvent the above problems, we use an approximation to the entropy, namely, the varimax norm. The varimax norm of an image is defined as <sup>13</sup>

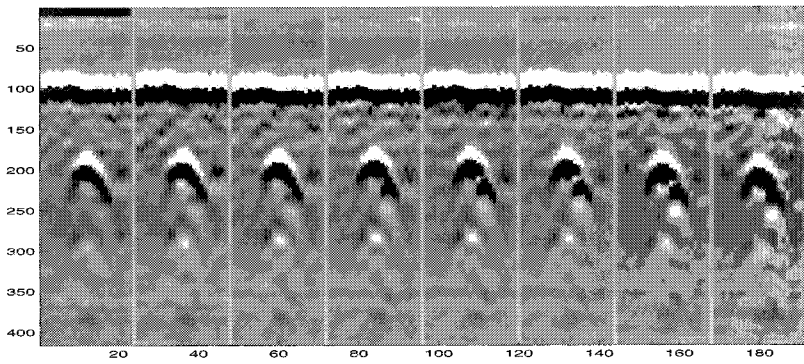
$$\Omega(X) = \frac{[\sum_m^M \sum_n^N X^2(m, n)]^2}{\sum_m^M \sum_n^N X^4(m, n)} \quad (7)$$

An image of small entropy has a small varimax norm and vice versa. An identically homogeneous image, i.e., a constant image, has an entropy of one, which is the smallest varimax norm. Any inhomogeneous image will have a varimax norm larger than one. In entropy optimized CS, we start with a small lower boundary  $a$  and a

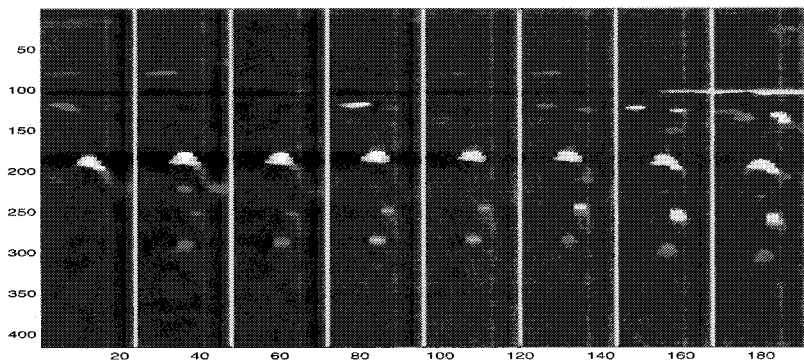




(a)

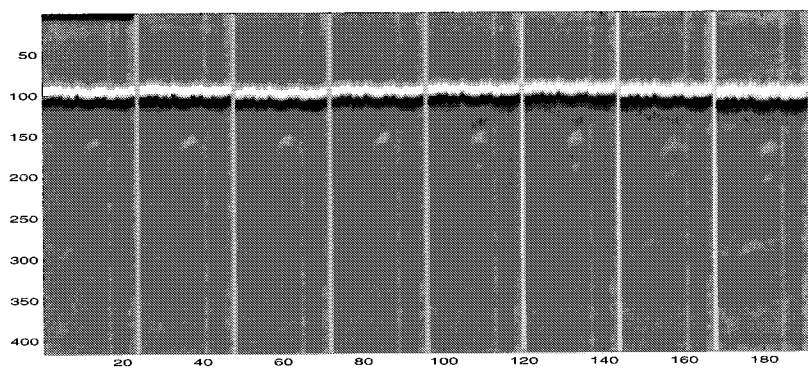


(b)

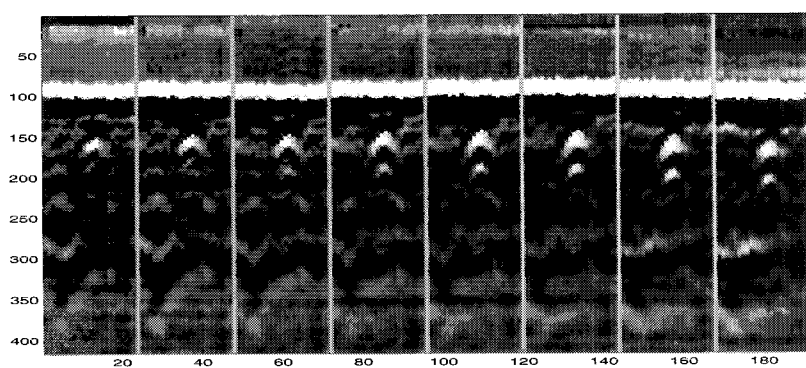


(c)

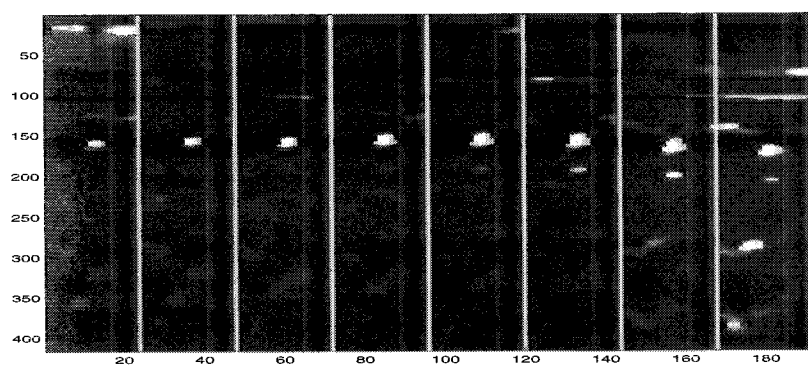
Fig. 9. (a) Cascaded GPR images over an M15 landmine, (b) cascaded images of the ICS at iteration 1, (c) cascaded images of the ICS at iteration 6.



(a)



(b)



(c)

Fig. 10. (a) Cascaded GPR images over a VS-2.2 landmine, (b) cascaded images of the ICS at iteration 1, (c) cascaded images of the ICS at iteration 6.

large upper boundary  $b$ . Then we increase  $a$  and decrease  $b$  at the same rate and calculate the varimax norm of the resulting image. When a minimum is reached, it is believed that at least a sub-optimal set of boundaries has been encountered and the above process stops.

```

(1) Initialize  $a_0$  and  $b_0$ 
(2) Choose step size  $\mu$ 
(3) FOR  $k = 1 : K$ 
    (a)  $a_k = a_0 + \mu k, b_k = b_0 - \mu k$ 
    (b) Contrast stretch image  $X$  by
            
$$Y = \text{CS}(X, a_k, b_k)$$

    (c) Compute the varimax norm of  $Y$ 
            
$$v(k) = \Omega(Y)$$

    (d) IF  $v(k - 2) \geq v(k - 1)$  AND  $v(k - 1) \leq v(k)$ 
        
$$Y_{opt} = \text{CS}(X, a_k, b_k)$$

        STOP
    ENDIF
(4) ENDFOR

```

Fig. 11. Entropy optimized contrast stretch.

In theory  $a$  and  $b$  can be found by exhaustive search through all the possible combinations. In practice we use entropy optimized CS to find  $a$  and  $b$  in a sub-optimal way. As the examples show, the results are, nevertheless, very good and make detection easy. Therefore we can conclude that the sub-optimal searching does not compromise the final results as landmine detection is concerned. Details of the entropy optimized CS is shown in Fig. 11. To find the optimal boundaries corresponding to the minimum entropy, we need to find the point  $k$  such that varimax norm  $v(k - 1) \leq v(k) \leq v(k + 1)$ .

#### 4.2. Examples

Using field data, we demonstrate that our method is able to produce results of high quality, i.e., the object reflected signals are enhanced. Fig. 12(a) and (b) display a raw GPR image and the resulting image of subtracting its columnwise ensemble average. In both images, it is difficult to observe the object, a buried landmine. As a way of seeing how difficult to detect the object using these two images, we plot the corresponding outputs of the energy detector in Fig. 13(a) and (b). It is seen that the energy detector cannot make the correct decision. In Fig. 13(c) to (h) we display the resulting images of entropy optimized CS. Because the original GPR image takes discrete values between 0 and 1.0, we start with a lower boundary  $a_0 = 0.01$  and

upper boundary  $b_0 = 0.99$  and a step size  $\mu = 0.01$ . The lower boundary is increased as

$$a_k = a_0 + \mu k \quad (8)$$

where  $k = 1, \dots, 20$ . Similarly the upper boundary is decreased as

$$b_k = b_0 - \mu k. \quad (9)$$

As the two boundaries close in, the object reflected signals, i.e., the hyperbolic curve in the image, becomes clearer and details begin to come forward. On the other hand, when the two boundaries get too close to each other, too much details of the GPR image is brought up and actually affect the detection. The resulting image with the smallest varimax norm is obtained at  $k = 16$ , Fig. 12(f). Fig. 13(c) to (h) show the outputs of the energy detector. For comparison, all the outputs are normalized to one and we can see that Fig. 13(f) has the lowest noise level than all the other plots. For larger  $k$ , the output of the energy detector becomes worse. Changes in the varimax norm is plotted in Fig. 14. The curve has a bowl-shape and the minimum happens at  $k = 13$ . From Fig. 12 and 13 we see that at  $k = 13$  we obtain both a homogeneous image and lowest noise level in the output of the energy detector. Another examples is shown in Fig. 15. The GPR image is taken over a buried landmine TMA4. Because of the dominance of the specular reflection, the landmine reflected signal is not observable either in the original image or the image after subtracting the mean. Using CS with entropy regularization, the optimal resulting image is shown in Fig. 15(c). The landmine reflected signals can be easily detected, Fig. 15(d).

## 5. Conclusions

Landmine detection and localization using GPR is an active research area. Many methods have been developed to remove clutters in GPR image and to obtain better detection and localization performance. In this chapter we introduce two methods based on histogram modification, which is a popular image enhancement technique. As we have shown, using a proper technique can greatly improve the quality of GPR image. In some cases, enhanced GPR image makes landmine detection an easy and straightforward task. In other cases, post-processing are needed to obtain better results.

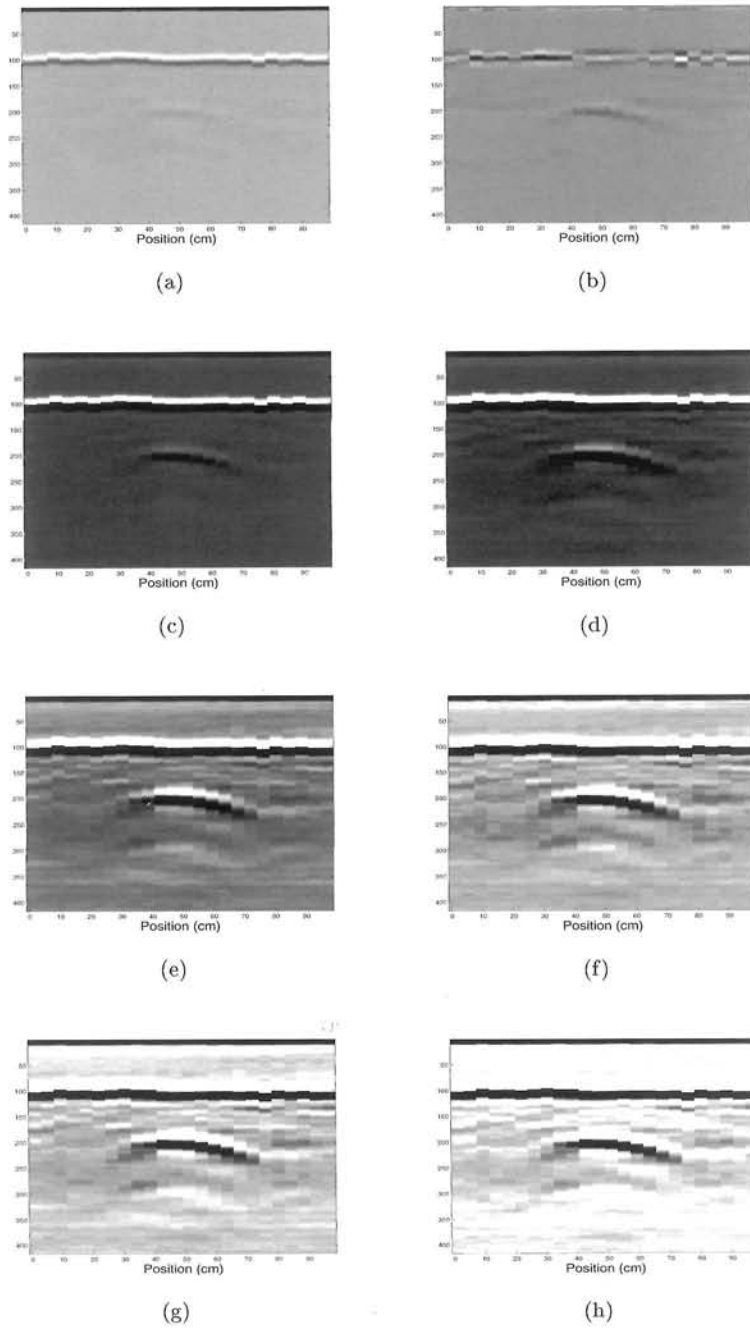


Fig. 12. (a) Observed GPR image, (b) subtracted by the ensemble average, (c) - (h), resulting images of CS at  $k = 7, 9, 11, 13, 15, 17$ .

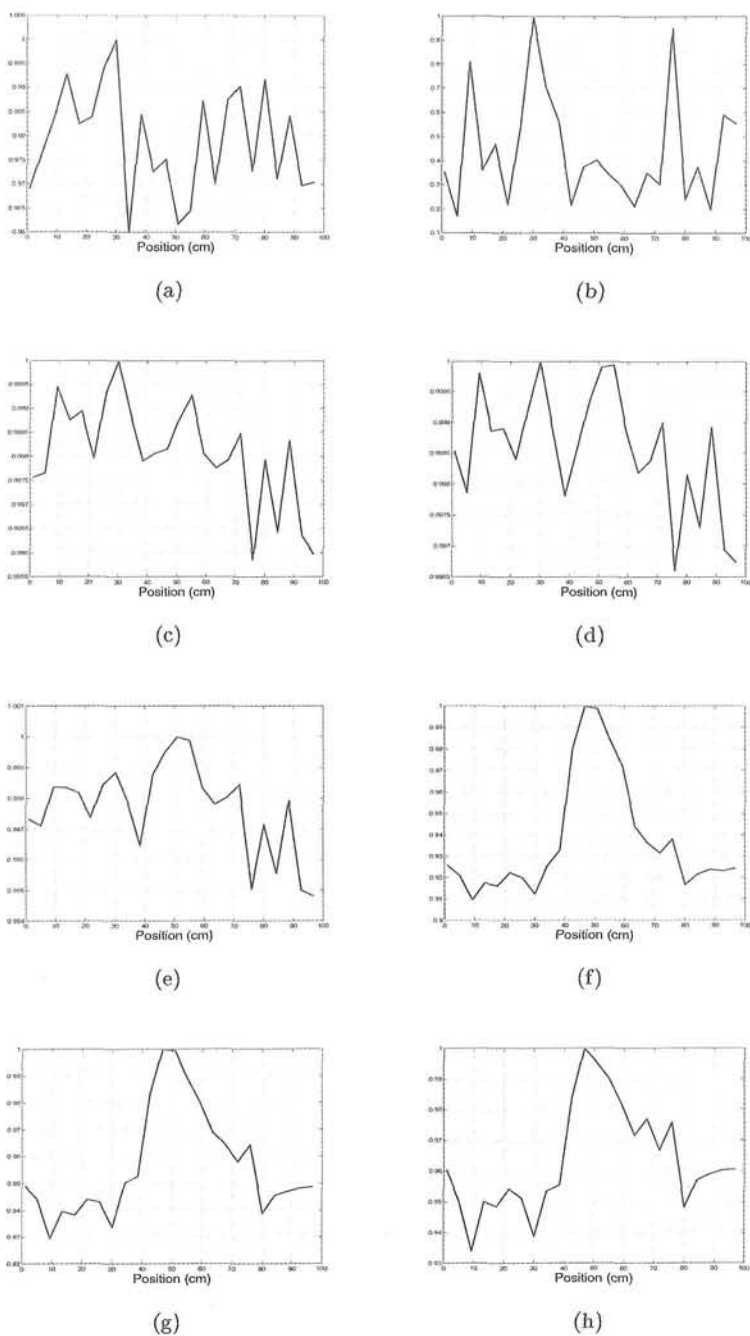


Fig. 13. Output of the energy detector from (a) the observed GPR image, (b) the GPR image subtracting the ensemble average, (c)-(h) the resulting images of CS at  $k = 7, 9, 11, 13, 15, 17$ .

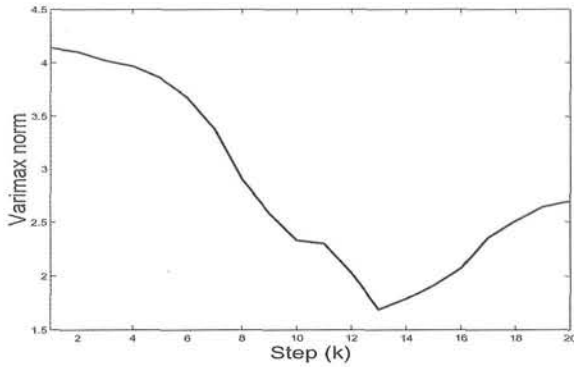


Fig. 14. Varimax norm versus lower boundary  $a$ .

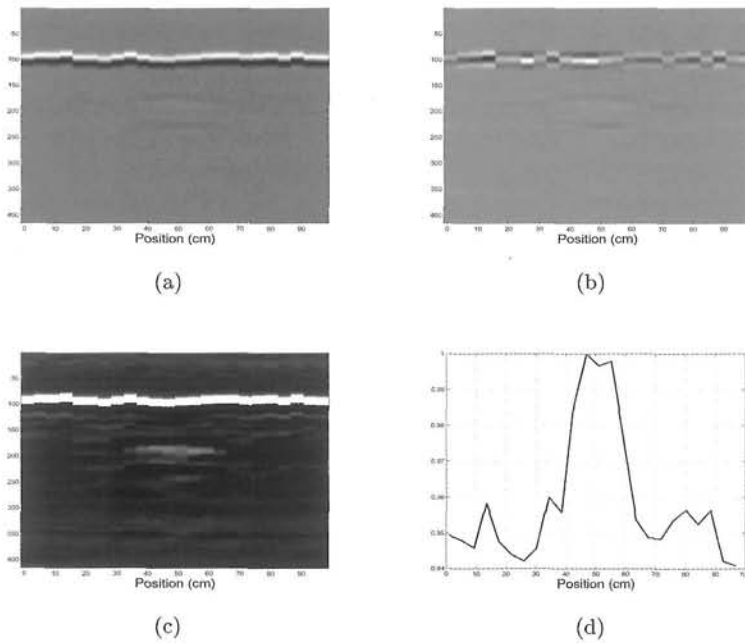


Fig. 15. (a) Observed GPR image, (b) subtracted by the ensemble average, (c) resulting images of CS at optimal stage, (d) output of the energy detector of (c).

## References

1. L. Peters Jr., J.J. Daniels, and J.D. Young, "Ground penetrating radar as a subsurface environmental sensing tool", *Proc. IEEE*, vol. 82, no. 12, pp. 1802–1822, Dec. 1994.
2. A.J. Witten, J.E. Molyneux, and J.E. Nyquist, "Ground penetrating radar tomography: algorithms and case studies", *IEEE Trans. Geosci. Remote Sensing*, vol. 32, no. 2, pp. 461–467, Mar. 1994.

3. M.S. Uppsall, L.M. Pettersson, M. Georgson, and S. Sjokvist, "Temporal IR contrast variation of buried landmines", in *Proc. SPIE AeroSense Detection and Remediation Technologies for Mines and Minelike Targets V*, Orlando, FL, Apr. 2000, vol. 4038, pp. 146–155.
4. I.K. Sendur and B.A. Baertlein, "Numerical simulation of thermal signatures of buried mines over a diurnal cycle", in *Proc. SPIE AeroSense Detection and Remediation Technologies for Mines and Minelike Targets V*, Orlando, FL, Apr. 2000, vol. 4038, pp. 156–167.
5. J.L. Drewniak and W. Cui, "Detection of electronic mines, timers, and fuses through electromagnetic interference signatures and stimulated emissions", in *Proc. SPIE AeroSense Detection and Remediation Technologies for Mines and Minelike Targets V*, Orlando, FL, Apr. 2000, vol. 4038, pp. 66–77.
6. M. Ozdemir, E.L. Miller, and A.J. Witten, "Clutter modeling and estimation methods for low metal content mine characterization from broadband electromagnetic induction data", in *Proc. SPIE AeroSense Detection and Remediation Technologies for Mines and Minelike Targets V*, Orlando, FL, Apr. 2000, vol. 4038, pp. 45–55.
7. S.W. McKnight, C.A. DiMarzio, and W. Li, "Laser-induced acoustic generation for buried object detection", in *Proc. SPIE AeroSense Detection and Remediation Technologies for Mines and Minelike Targets V*, Orlando, FL, Apr. 2000, vol. 4038, pp. 734–739.
8. C.A. DiMarzio, W. Li, L.J. Berg, and J.M. Sabatier, "Toward a laser-based, non-contact acoustic landmine imager", in *Proc. SPIE AeroSense Detection and Remediation Technologies for Mines and Minelike Targets V*, Orlando, FL, Apr. 2000, vol. 4038, pp. 740–747.
9. A.D. Hibbs et al, "Field tests results of a nuclear quadrupole resonance landmine detection system", in *Proc. SPIE AeroSense Detection and Remediation Technologies for Mines and Minelike Targets V*, Orlando, FL, Apr. 2000, vol. 4038, pp. 564–571.
10. D. Daniels, *Surface-penetrating Radar*, IEE, London, 1996.
11. C.M. Rappaport, S.G. Azevedo, T. Rosenbury, and J. Gough, "Handheld forward-looking focused array mine detection with plane wave excitation", in *Proc. SPIE AeroSense Detection and Remediation Technologies for Mines and Minelike Targets V*, Orlando, FL, Apr. 2000, vol. 4038, pp. 1118–1126.
12. A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
13. H. Wu and J. Barba, "Minimum entropy restoration of star field images", *IEEE Trans. Systems, Man, and Cybernetics*, vol. 28, no. 2, pp. 227–231, Apr. 1998.



This page is intentionally left blank

## CHAPTER 25

### INFRA-RED IMAGE PROCESSING

Margaret J. Varga and Paul Ducksbury

*QinetiQ, St. Andrew's Road, Malvern, Worcestershire, WR14 3PS, United Kingdom*

*E-mail: m.varga@signal.qinetiq.com*

© Copyright QinetiQ Ltd 2003

**Abstract:** Infra-red (IR) technology is applied in a wide range of application domains, e.g. military, medical, security and others. All objects, live or dead and of any colour, emit infra-red radiation by virtue of their temperature; the exact degree of radiation is determined by the absolute temperature and the thermal characteristics of the material from which it is made. The radiation is present day or night with or without external illumination. Infra-red technology is concerned with the detection and imaging of this emitted infra-red radiation. It helps to visualise objects that cannot be seen by the naked eye. Infra-red imaging is therefore a method for producing an image of the heat emitted from any object's surfaces. A thermogram is a calibrated graphic record of the temperature distribution obtained by thermography.

#### 1. Introduction

Infra-red (IR) technology is applied in a wide range of application domains, e.g. military, medical, security and others. All objects, live or dead and of any colour, emit infra-red radiation by virtue of their temperature; the exact degree of radiation is determined by the absolute temperature and the thermal characteristics of the material from which it is made. The radiation is present day or night with or without external illumination. Infra-red technology is concerned with the detection and imaging of this emitted infra-red radiation. It helps to visualise objects that cannot be seen by the naked eye. Infra-red imaging is therefore a method for producing an image of the heat emitted from any object's surfaces. A thermogram is a calibrated graphic record of the temperature distribution obtained by thermography.

In this chapter, both the hardware and some of the processing techniques associated with infra-red technology are discussed. Examples of military and medical applications will demonstrate the versatility of infra-red imagery.

##### *1.1 Infra-red Wavebands*

Spectrally infra-red radiation is located between the visible and radio frequencies. Infra-red radiation is generally thought of in three spectral bands: Short Wavelength Infra-red (SWIR) also called near infra-red, lying between 0.7 - 2.0 $\mu\text{m}$ ; Medium Wavelength Infra-red (MWIR) ranging from 3.0 $\mu\text{m}$  to 5.0 $\mu\text{m}$ ; and Long Wavelength Infra-red (LWIR) between 8.0 - 14.0 $\mu\text{m}$ . Both the MWIR and the LWIR are strongly absorbed by water and organic compounds. Infra-red sources are either thermal (i.e. emitted by matter in the temperature range 100-

3000 K) or electronic (i.e. emitted by high-energy electrons interacting with magnetic fields) [1-4].

## ***1.2 Infra-red Detectors***

There are two types of infra-red detectors, those that require cooling (cooled) and those that do not (uncooled). Cooled infra-red detector systems are bigger and more expensive than uncooled detector systems but are more sensitive and can produce higher resolution images. Uncooled detectors on the other hand are cheaper, lighter, more compact and hence more portable; at present, however, they tend to be less sensitive and are commonly used for 8-14 $\mu$ m only.

The heart of a cooled thermographic camera is an infra-red photo-detector typically made of indium antimonide (InSb) or mercury cadmium telluride (MCT). InSb allows larger more uniform arrays at present, while MCT allows access to the long and short wavelength bands and also allows higher operating temperatures. The detector lies in the focal plane of the camera and is cooled by liquid nitrogen. Uncooled infra-red detectors use pyroelectric bolometer techniques. The infra-red radiation from an object is focused on the detector - the focusing system can be based on either refractive or reflective optics. The thermogram can be produced by an array of detectors which converts the infra-red radiation directly into an image. The signal can be represented as a grey-level or colour coded image [5].

Alternatively, a scanning system can be used in which an image is built up by mechanically scanning the scene onto a single detector or a linear or two-dimensional array. Indeed the generation II systems based on scanned linear arrays of photodiodes, have very high spatial resolution with good thermal sensitivity and are available commercially. An example of such system is the UK STAIR 'C' system which has an array of 768x8 LWIR MCT detectors [6]. The system produces an image of 1280x768 pixels with a noise equivalent temperature difference (NETD) of 50mK.

Long linear array detectors can be used with simple scanning mechanisms to generate high performance imaging over wide fields of view. Two-dimensional focal plane arrays increasingly provide the basis of systems which require no scanning and offer high sensitivity. There is a wide and developing range of infra-red focal plane array sensors using different detector technologies. Common to the development of all these arrays is the continual increase in thermal sensitivity. This enhanced sensitivity may be used directly or compromised to provide different operating designs. The available optical system and the required frame rate determine the choice of read-out process required to achieve a given sensitivity [7-9].

Recently medium wavelength infra-red arrays have been developed which have 1024\*768 pixels on a 26 micron pitch [6]. These arrays are made from epitaxially grown indium antimonide instead of the more conventional InSb. There are two advantages of this approach, namely the photodiodes can be grown

on degenerately doped InSb substrates which a high degree of transparency, so the requirement for the substrate to be thinned is much reduced, this means simplified manufacture. It also offers the potential for an increase in operating temperature of many tens of degrees, through elimination of contact leakage currents. At present the operation of these devices has been focussed on temperatures of 80K so that they can be compared with the conventional structures. Initial results have shown that the arrays have high operability, despite the need to stitch reticules in the fabrication of the silicon read-out circuit and the temperature sensitivity, which is close to the theoretical limit. Furthermore, imaging from the arrays has compared very favourably with those taken using generation II cameras and it is likely that this approach could offer a cost effective way to large format MWIR system. Examples of images are shown in Figures 1 and 2. In the case of Figure 1, which was taken indoors, the depth of field was  $f/4.5$  and the integration time was 3ms. In Figure 2 an outdoor scene was captured during an early December evening, the depth of field was  $f/2.3$  and the integration time was 1.8ms.



*Fig. 1: An Indoors Image from 1024 x 768 InSb arrays at 77K*



*Fig. 2: An Outdoors Image from 1024 x 768 InSb arrays at 77K*

## **2. Military Infra-red Imaging**

Infra-red technology is applied in a variety of military applications and there is a need for both cooled and un-cooled systems. Cooled IR detectors offer high

performance when required, for example in; weapon sights, surveillance systems, remote ground sensors, imaging IR seekers, non-co-operative target recognition, mine sensors, driving aids, fire fighting and rescue. Where lower performance is acceptable, un-cooled infra-red detectors reduce the logistic burden and their low power and compactness is particularly useful for highly portable, remote and autonomous operation.

### 3. Military Surveillance: Downward Looking IR Imagery

The fundamental task here is generically defined as the location of some region of interest in an image, for example, the fast and automatic detection of urban regions. This could be used as a cueing aid for more detailed processing, such as the detection of road networks or junctions and buildings in order to allow registration of imagery to maps for navigation, for change and target detection, image distortion correction as well as map-update. It could also be an attention cueing device for human image interpreters.

This section describes the use of the Pearl Bayes Networks (PBN's) for the automatic extraction of knowledge about regions from infra-red linescan imagery, i.e. surveillance imagery.

#### 3.1 Infra-red Linescan Imaging and Correction

The aerial infra-red linescan imagery used in this application is produced by a sensor which has a single detector. Scanning in the x direction is achieved by via a rotating mirror which has a uniform angular velocity and gives a 120° field of view. Scanning in the y-direction is achieved by the aircraft motion. The scanner arc introduces a characteristic ( $\text{Sec}^2$ ) distortion into the imagery at either extreme of the 120°arc. This can be corrected with a relatively simple trigonometric transformation. Figure 3 illustrates the relationship in which  $h$  is the height of the aircraft,  $\beta$  is the bank angle of the aircraft,  $\Delta x$  is the ground resolution of a single pixel with  $\Delta\theta$  being the corresponding swathe angle for that pixel. The equation for the correction is  $\Delta x = \Delta\theta h / \cos^2(\theta + \beta)$ .

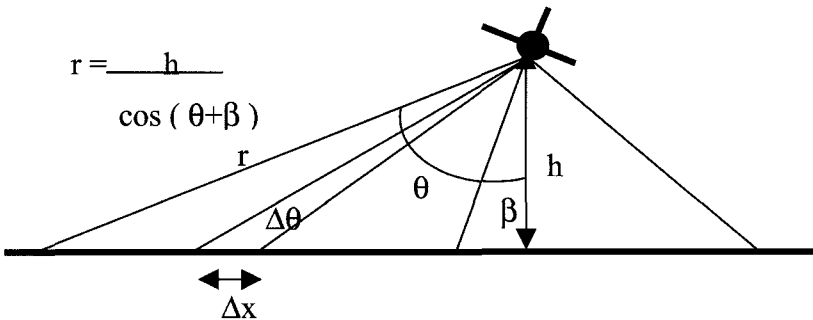


Fig. 3: Linescan Sensor Distortion

### 3.2 Pearl's Bayes Networks (PNB)

Pearl's Bayes networks [10] are directed acyclic graphs, see Figure 4. In this graph nodes B, C and E represent different statistical information extracted from the image, whilst node A represents the 'belief' in detecting, say, an urban patch in an image. A graph G is a pair of sets  $(V, A)$  for which  $V$  is non-empty. The elements of  $V$  are vertices (nodes) and the elements of  $A$  are pairs  $(x,y)$  called arcs (links) with  $x \in V$  and  $y \in V$ .

Consider the simple network that is shown in Figure 4. Here the symbol  $\pi$  represents the causal support (or evidence) whilst  $\lambda$  represents the diagnostic support (or evidence).  $G_{BA}^+$  and  $G_{BA}^-$  are sub-graphs as described in the next section, equations for computing belief and propagation of information are given in the following sections<sup>1</sup>.

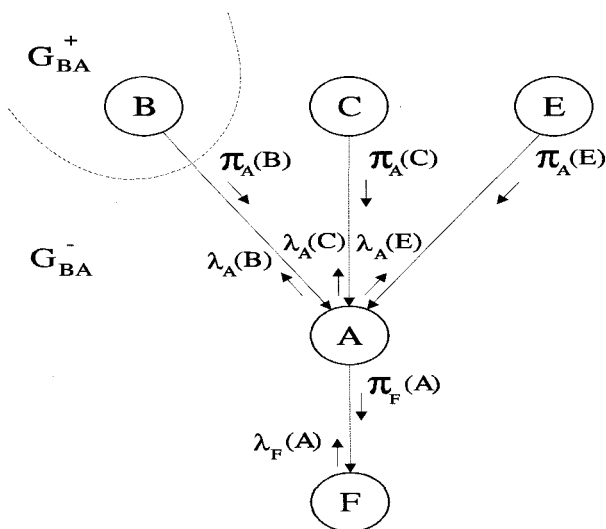


Fig. 4: Network

#### 3.2.1 Belief Equations

Consider the link from node B to A then the graph G consists of the two subgraphs  $G_{BA}^+$  and  $G_{BA}^-$ . These two subgraphs contain the datasets  $D_{BA}^+$  and  $D_{BA}^-$  respectively.

From Figure 4 it can be observed that node A separates the two subgraphs  $G_{BA}^+ \cup G_{CA}^+ \cup G_{EA}^+$  and  $G_{AF}^-$ . Given this fact we can write the equation:

$$P(D_{AF}^- | A_i, D_{BA}^+, D_{CA}^+, D_{EA}^+) = P(D_{AF}^- | A_i) \tag{1}$$

by using Bayes rule the belief in  $A_i$  can be defined as:

---

<sup>1</sup> The equations are derived along similar lines to those derived by Pearl [10] where in his example node A has just two predecessors and two successors.

$$\begin{aligned}
BEL(A_i) &= P(A_i | D_{BA}^+, D_{CA}^+, D_{EA}^+, D_{AF}^-) \\
&= \alpha \cdot P(A_i | D_{BA}^+, D_{CA}^+, D_{EA}^+) \cdot P(D_{AF}^- | A_i) \\
&= \alpha \cdot P(D_{AF}^- | A_i) \cdot \left[ \sum_{j,k,l} P(A_i | B_j, C_k, E_l) \cdot P(B_j | D_{BA}^+) \cdot P(C_k | D_{CA}^+) \cdot P(E_l | D_{EA}^+) \right]
\end{aligned} \tag{2}$$

where  $\alpha$  is taken to be a normalising constant and  $i, j, k$  and  $l$  range of the number of variables in A, B, C and E respectively. It can be seen Eq. (2) is computed using three types of information:

- Causal support  $\pi$  (*from the incoming links*).
- Diagnostic support  $\lambda$  (*from the outgoing links*).
- A fixed conditional probability matrix (*which relates A with its immediate causes B, C and E*).

Firstly, the causal support equations:

$$\pi_A(B_j) = P(B_j | D_{BA}^+) \tag{3}$$

$$\pi_A(C_k) = P(C_k | D_{CA}^+) \tag{4}$$

$$\pi_A(E_l) = P(E_l | D_{EA}^+) \tag{5}$$

Secondly, the diagnostic support equation is given by

$$\lambda_F(A_i) = P(D_{AF}^- | A_i) \tag{6}$$

Finally, the conditional probability matrix is defined to be

$$P(A|B, C, E) \tag{7}$$

The belief Eq. (2) can now be rewritten to obtain the belief at node A based on the observations at B, C and E, e.g. the belief that an urban region is detected:

$$BEL(A_i) = \alpha \lambda_F(A_i) \cdot \sum_{j,k,l} P(A_i | B_j, C_k, E_l) \cdot \pi_A(B_j) \cdot \pi_A(C_k) \cdot \pi_A(E_l) \tag{8}$$

The belief at nodes B, C and E can be obtained from the equations:

$$BEL(B_j) = \alpha \pi_A(B_j) \cdot \lambda_A(B_j) \tag{9}$$

$$BEL(C_k) = \alpha \pi_A(C_k) \cdot \lambda_A(C_k) \tag{10}$$

$$BEL(E_l) = \alpha \pi_A(E_l) \cdot \lambda_A(E_l) \tag{11}$$

In other words the belief is the resultant product of causal support information, diagnostic support information (*belief*) and prior knowledge. The propagation equations described below are iterated to support belief of a certain event.

### 3.2.2 Propagation Equations

The propagation equations for the network are derived as follows, firstly the diagnostic support. From a previous analogy with Eq. (6) we can write

$$\lambda_A(B_i) = P(D_{BA}^- | B_i) \tag{12}$$

by partitioning  $D_{BA}^-$  into its component parts, namely  $A$ ,  $D_{AF}^-$ ,  $D_{CA}^+$ ,  $D_{EA}^+$  we can obtain

$$\lambda_A(B_i) = \alpha \sum_{j,k} \left[ \pi_A(C_j) \cdot \pi_A(E_k) \cdot \sum_l \lambda_F(A_l) \cdot P(A_l | B_i, C_j, E_k) \right] \quad (13)$$

likewise for  $\lambda_A(C_j)$  and  $\lambda_A(E_k)$

$$\lambda_A(C_j) = \alpha \sum_{i,k} \left[ \pi_A(B_i) \cdot \pi_A(E_k) \cdot \sum_l \lambda_F(A_l) \cdot P(A_l | B_i, C_j, E_k) \right] \quad (14)$$

and

$$\lambda_A(E_k) = \alpha \sum_{i,j} \left[ \pi_A(B_i) \cdot \pi_A(C_j) \cdot \sum_l \lambda_F(A_l) \cdot P(A_l | B_i, C_j, E_k) \right] \quad (15)$$

### 3.2.3 Causal Equations

These are defined using a similar analogy as follows.

$$\pi_F(A_i) = P(A_i | D_{BA}^+, D_{CA}^+, D_{EA}^+) \quad (16)$$

and from this we then derive the equation

$$\pi_F(A_i) = \alpha \left[ \sum_{j,k,l} P(A_l | B_i, C_j, E_k) \cdot \pi_A(B_j) \cdot \pi_A(C_k) \cdot \pi_A(E_l) \right] \quad (17)$$

An important point to realise from these Eqs. (13 -15 & 17) is the fact that they demonstrate that the parameters  $\lambda$  and  $\pi$  are orthogonal to each other, i.e. perturbation of one will not affect the other. Hence evidence propagates through a network and there is no reflection at boundaries.

### 3.3 Region Segmentation Using Pearl's Bayes Networks

The above Pearl's Belief network approach [10] has been adapted for the detection of urban regions [11] using a high powered parallel processing system for improved performance. The belief network is used in a multi-resolution sense to combine statistical measures of texture into the detection (or belief) of the required region<sup>2</sup> see Figure 5. The problem is approached by taking several statistical measures from small patches of an image which are treated as a set of judgements about the content of the patches. These statistics are the number of edges, the number of extrema and grey level distribution type. These statistics are quantised down into a smaller number of levels. The number of edges and extrema are both reduced to five levels, whilst the distribution type has four possibilities. It is important to stress that any suitable measure that provides the required textural or thermal discrimination could have been used.

The statistics are then used to produce a set of judgements; for example, an expert might, upon looking at a particular window, issue a report of the form (0.0,0.7,0.9,0.6,0.0). This means that he believes there is a 70 % chance that

---

<sup>2</sup> The authors have also carried out work into using the Belief Network approach at a higher level of abstraction, i.e. for combination of several region finding algorithms.



level 2 describes the number of edges, 90% chance that its level 3 will do the same and 60% for level 4. But he believes there to be no chance of it being levels 1 or 5.

S1 and S3: Edges, extrema, distribution type  
From fine resolution image  
S4-S6: Edges, extrema, distribution type  
From coarse resolution image  
Bf: Belief from fine resolution  
Bc: Belief from coarse resolution  
B: Belief

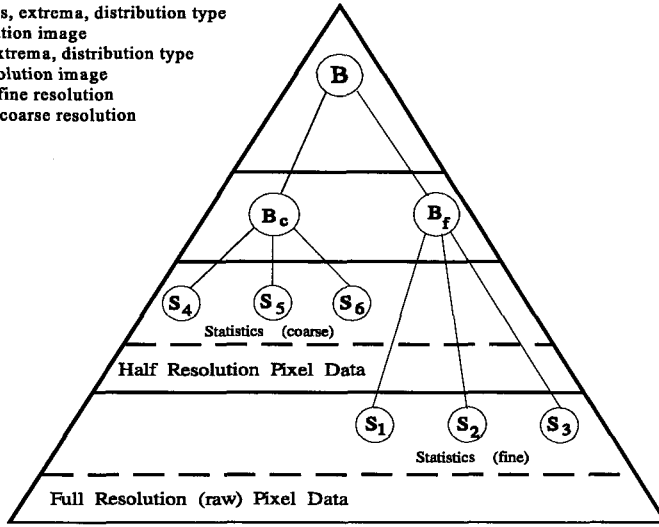


Fig. 5: Multi-resolution Approach

For the purpose of the system described here the Belief at nodes  $B_f$ ,  $B_c$  and  $B$  in Figure 3 is quantised to three levels, namely (*low, medium, high*).

The fixed conditional probability matrices (*i.e.*  $P(B_f|s1,s2,s3)$  etc.), which are the prior information and relate the given node with its causal information, are created along similar lines to the approach used in [12] and [13]. They are based upon the assumption that the probability of an event at a given node should be greater if its causal information is more tightly clustered together than it should be if the causal information is further apart. For the  $P(B|B_f,B_c)$  matrix (*which relates the beliefs from the fine and coarse resolutions*) slightly more emphasis is given to the causal information received from the coarse resolution belief.

$P(B_f|s1_j,s2_k,s3_l)$  is described formally as

$$P(B_f|s1_j,s2_k,s3_l) = \begin{cases} 0.75 & \text{if } i = j = k = l \\ 0.25/\alpha & \text{if } (i \neq j = k = l) \wedge (0 < |i - j| \leq C) \\ 1.0/\beta & \text{if } \neg(j = k = l) \\ & \wedge (\max(j,k,l) - \min(j,k,l) \leq 2C) \\ & \wedge (\min(j,k,l) \leq i \leq \max(j,k,l)) \\ 0.0 & \text{otherwise} \end{cases} \quad (18)$$

such that  $\sum_{j,k,l} P(B_i | s1_j, s2_k, s3_l) \leq 1 \forall i$  where  $C = 1$  and  $i, j, k, l$  range over the number of variables in  $B_i$ ,  $s1$ ,  $s2$  and  $s3$  respectively.  $\alpha$  and  $\beta$  represent the number of different values of  $i$  satisfying the constraint.

$P(B_i | B_f, B_c)$  is defined as

$$P(B_i | B_f, B_c) = \begin{cases} 0.9 & \text{if } i = j = k \\ 0.7 & \text{if } (i = j) \wedge (|i - k| \leq 1) \\ 0.3 & \text{if } (i = k) \wedge (|i - j| \leq 1) \\ 0.6 & \text{if } (i = j) \wedge (|i - k| > 1) \\ 0.1 & \text{if } (i = k) \wedge (|i - j| > 1) \\ 0.0 & \text{otherwise} \end{cases} \quad (19)$$

such that  $\sum_{j,k} P(B_i | B_f, B_c) \leq 1, \forall i$  where  $i, j, k$  range over the number of variables in  $B$ ,  $B_f$  and  $B_c$  respectively.

### 3.4 Performance

An example result is shown in Figure 6, this demonstrates the location of urban regions in a near infra-red image. The image was taken during the night at approximately 4000 feet. An outline has been drawn around the areas labelled as most likely to be urban. It is possible to post-process such a result to remove small isolated regions, which are likely to be erroneous classifications, this has not been done here. A probability surface could also have been integrated into the image to produce a smoother transition between the areas of different classification.



Fig. 6: Outline of urban region

This system can be easily adapted to alternative applications of a broadly similar nature, i.e. classifying or clustering regions. The only change necessary may be a different set of statistics which more accurately describe the detail required in the image. In addition, if the number of input nodes alter, then the prior knowledge in the fixed conditional probability matrix will need to change, however, the set of basic equations given previously can be used to automatically generate this information.

This approach has been demonstrated successfully for the texture segmentation of driveable regions for autonomous land vehicles, and for urban region segmentation in both SPOT, and Russian Satellite imagery in [11,14 and 15].

## 4. Target Detection and Tracking in Forward looking IR Imagery

### 4.1 Introduction

One of the most powerful features in any modern battlefield surveillance system is the ability to carry out automatic detection and tracking of targets. The amount of information being presented to human operators is increasing at an alarming rate and needs to be reduced. Any system that can simply filter the data and present filtered/screened results of possible targets will be of significant benefit.

The task here is generically defined as the detection of small point sized ‘hot’ targets in wide area IR surveillance imagery. These potential targets will be within some operator specified sizes, typically  $2 \times 2$  to  $10 \times 10$  pixels. The resulting detection can be used to aid subsequent target recognition through cueing for a narrow field of view imager and/or a human operator. The requirement for the system described here was not necessarily to locate all targets or indeed just the targets, but rather to locate possible areas of interest for further analysis. The wide field of view sensor used produces very small potential targets. Furthermore, these targets have low contrast and low signal-to-noise ratios which makes their detection difficult.

### 4.2 System Overview

The system combines some ‘conventional’ image processing techniques with morphological analysis to perform automatic cueing and tracking of small objects/targets. Most stages of the process have been specifically chosen because of their suitability for implementation in special hardware modules. The process is shown schematically in Figure 7. Only the main elements of this system are considered below.

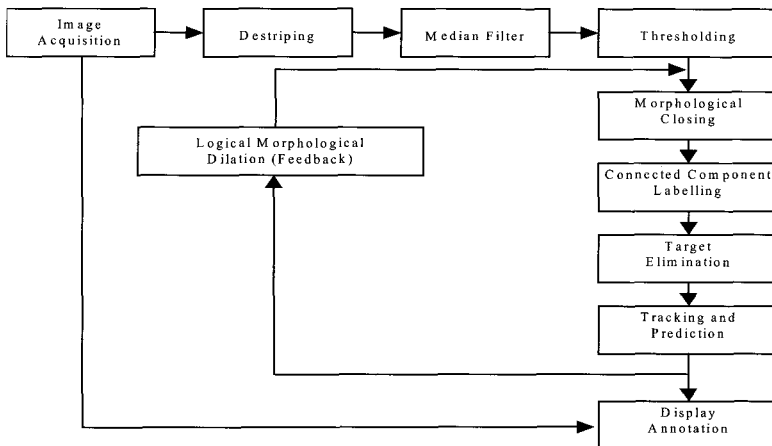


Fig. 7: Algorithm Block Diagram

### 4.2.1 Destriping

Sensor imperfections mean that the imagery contains marked fixed pattern stripe artifacts. A destriping stage is therefore required to reduce or even to eliminate the banding effects that occur in this type of imagery. It is likely that future generations of infra-red imagers for this application will produce images of improved quality with imperceptible striping artifacts and hence such a destriping algorithm will become unnecessary.

The destriping algorithm [16] removes bias variation between adjacent scan lines. Two adjacent scan lines  $i$  and  $i+1$  are equalised by determining the distribution differences between adjacent pixels along the scans. The majority of entries should reflect differences in the baseline between the two scan lines. An additive correction to scan line  $i+1$  is obtained from the median of the difference distribution, the median being used as it is a statistically robust measurement. This process is then repeated using scan lines  $i+1$  and  $i+2$  and so on.

The disadvantage of this approach is that the software implementation is relatively slow. However, an alternative scheme was developed as an intermediate measure (*prior to hardware design and implementation of the full destriper*). The approach is to model the sensor responses in order to estimate a set of corrective factors. If it is assumed that the image is uniform over a number of neighbouring scan lines then any differences should be due to the sensor itself. The median of each image row is obtained and the maximum of all medians is taken as the maximum sensor response. The difference of the maximum response and the median of each row can then be used as an additive amount for that row. The analysis can be done on the first frame in the sequence and then at successive frames the corrective (held as a look up table) amounts are simply added to each row.

### 4.2.2 Median Filter

The median filter [17, 18] is used to remove salt-and-pepper type noise. In a small window over an image the pixel values are likely to be homogeneous with only a small number of them being attributable to noise. These noisy pixels tend to be at either extreme of the grey level distribution and therefore are unlikely to be selected as the output from a median filter (*the output being the median of the ranked input values*). This filter has the advantage of reducing noise without having a smoothing effect on the image. In this instance, since the possible targets are very small, only a  $2 \times 2$  median filter<sup>3</sup> is applied; despite this small window size the filter is successful in removing as noise.

---

<sup>3</sup> An important point to be considered is that a median filter when used for noise suppression could be replaced by grey scale morphology. This point does not really apply to this particular algorithm as the filter used only has a small kernel. It is, however, important and worth mentioning. A morphological opening followed by a closing operator can achieve the same effect as a median filter. Morphology has two distinct noise suppression stages, the opening suppresses positive noise impulses whilst the closing suppresses negative noise impulses.

### 4.2.3 Thresholding

Thresholding [18, 19] is used to segment objects from the background. It is useful, whenever possible, to calibrate detected feature values (*e.g. grey level*) so that a given amplitude interval represents a unique object characteristic. There are various useful adaptive thresholding schemes, for instance, based on the examination of local neighbourhood histograms or other measures. Unfortunately, these approaches tend to produce a significant amount of noise and in this application such noise would pose a major problem due to the small target sizes. In an effort to minimise the problem for this application the thresholding scheme used here is therefore based upon the global mean and variance of the portion of the imagery being processed. The threshold is set at  $\mu+3\sigma$ . This proved to be an acceptable level for this application domain, but an option has been provided in the algorithm for this to be varied interactively at run time (*see also the section below on local morphological dilation*).

### 4.2.4 Morphological Closing

The term morphology originally comes from the study of forms of plants and animals. In image processing it means the study of topology or structure of objects from their images. Morphological processing refers to certain operations where an object is 'hit' with a structuring element and hence reduced to a more revealing shape. Most morphological operations are defined in terms of two basic operations, namely erosion and dilation. Erosion is a shrinking operation, whereas dilation is an expansion operation; erosion of an object is accompanied by enlargement or dilation of the background. If  $X$  the object and  $K$  the morphological structuring element are thought of as sets in two-dimensional Euclidean space, then the erosion of  $X$  by  $K$  is the set of all points  $x$  such that  $K_x$  is included in  $X$ , where  $K_x$  is the transformation of  $K$  so that its origin is located at  $x$ . The dilation of  $X$  by  $K$  is the set of all points  $x$  such that  $K_x$  intersects at  $X$ . The morphological closing operator is defined as a dilation followed by an erosion.

Closing aims at blocking up narrow channels and thin lakes and is ideal for the study of inter-object distance. The reasons for applying a morphological closing operator in this application are twofold. Consider the thresholding of an image, this can and does result in the fragmentation of objects. Firstly, an object which is a target could be fragmented into several parts thus leading to the possibility of several targets being detected instead of one. Secondly, an object which is not a target (*perhaps by virtue of its size*) could be fragmented into small parts which are then likely to be identified as possible targets. To resolve this problem a morphological closing operator is applied in an attempt to piece the fragments back together. The kernel use is deliberately kept small to try to avoid merging several genuine targets.

A fuller and more detailed description of morphology can be found in numerous papers in the literature, see, for example [19-23].

#### 4.2.5 Connected component labelling

The objective of connected component labelling (CCL) is to take a binary image and apply a segmentation in order to obtain a set of connected regions, each of these disjoint regions being labelled with a unique identifier. Although this stage is currently performed in software a design and hardware module exists for future use [24]. This fundamental process is important in many applications and can be used as an input to shape recognition tasks.

#### 4.2.6 Target Elimination

Once a labelled image has been obtained regions can be discarded according to a number of criteria. The human operator will initially have specified a set of bounds for targets of interest. Regions which have a width, height or total area outside these constraints are discarded. It is not possible to discard objects based upon shape without knowledge of the type of targets which is unknown in this application domain.

#### 4.2.7 Tracking/Prediction

Once an acceptable set of regions has been obtained the co-ordinates of the centre points are passed to the tracking process. Tracking introduces an element of temporal consistency into the algorithm. This is used to resolve a number of issues such as false targets (*due to segmentation errors or genuine noise in the imagery*), targets appearing and disappearing, and overlapping targets. Once these issues have been resolved a prediction stage is performed to estimate the target's position in the next frame. Targets develop a "history" after  $n$  frames and therefore isolated noise which appears for  $n-1$  frames or less will not be tracked and can be eliminated.

The initial part of the tracking is actually an association stage where observations are associated with tracks. This uses a standard assignment optimisation algorithm [25] which was modified by [26] to deal with targets which appear and disappear. It was also modified by the authors to resolve the problem of several observations being identical distances from a given track but outside the permissible (gated) regions for all other tracks. This condition appeared to cause the standard algorithm to fail to converge to an optimum assignment.

Kalman filtering [27] is the classical approach for the prediction of a target's new position. It is the optimal predictor for tracking. If the  $x$  and  $y$  target co-ordinates can be decoupled then it has been shown [28,29] that the Kalman filter can reduce to the so called  $\alpha$ - $\beta$  filter which is much simpler and requires no matrix multiplication.

#### 4.2.8 Local Morphological Dilation (Adaptive Feedback)

An important point arising from the thresholding is the difficulty in setting a threshold level at just the correct level for detection of all targets. As has been

mentioned previously it is possible for noise to be included and also for a genuine target (*but maybe one which is very small and/or emitting low thermal radiation*) to be excluded from the thresholded image. The effect of this has been reduced by incorporating feedback from the tracking algorithm and essentially using the track confidence to adapt the output from the thresholded image. This is achieved simply by performing a local morphological dilation in an area around known targets (*targets that developed a history*). This attempts to enlarge the thresholded output to a point where it would be accepted by the target elimination stage and effectively reduces the number of drop-outs due to weak targets. If a target genuinely disappears then this approach will have no effect.

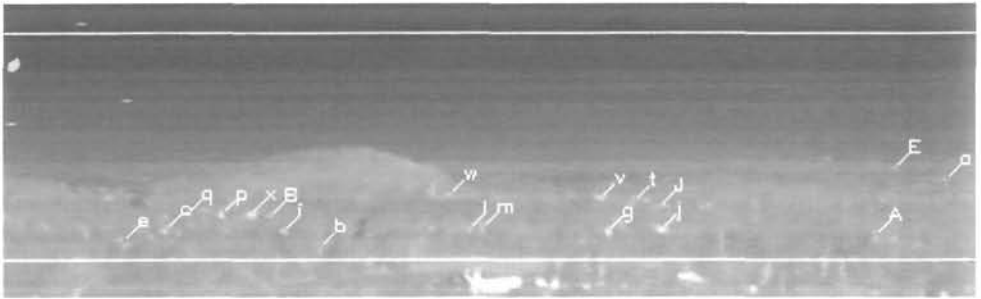


Fig. 8: Target detection

### 4.3 Performance

Targets (*or rather areas of possible interest*) are indicated by a diagonal line in Figure 8, the lower part of which points towards the targets centre whilst the upper part has a unique target identifier.

## 5. Medical Applications

### 5.1 Infra-red Imaging in Medicine

It was not until the early 1970s that an application of infra-red imaging was found in the field of medicine when military cameras were used as a means of examining the human body. Infra-red thermography has been used clinically since then. It provides a means of obtaining high quality images of thermal patterns on the surface of the human body. The IR systems used for medical applications should have a sensitivity that varies from 0.01°C to 1°C responding from 5°C to 40°C depending on the particular system and the part of the body to be examined [30]. Thermal images are observed in a greyscale or colour gradation. There is no defined colour scheme for the various temperature ranges; it is, however, common to use a spectrum of colours from darker to lighter as temperature increases. In these systems any error caused by variation in the detector's response can be calibrated out, for example, by alternatively detecting the radiation from the object and the radiation from a reference source.

The correlation between the skin temperature and underlying malignant disease was first realised in the case of breast cancer. This inevitably resulted in initial infra-red thermographic studies being concentrated on breast diseases [31]. However, there were problems, for instance, with limited sensitivity to deep lying tumours and poor control of environmental conditions during examination and recording.

The majority of the detected radiation comes from the topmost layers, i.e. body surface to 300 microns depth. The surface temperature of the skin is affected by both internal and external factors. The internal factors can be pathological or physiological, while the external factors are a function of ambient conditions, such as temperature, humidity and air flow [32]. Indeed ambient air flow is very important in medical thermography, and a uniform environment without any cooling draught, direct warmth of sunlight, or radiators etc. is essential. In general an ambient air temperature of between 18°C and 20°C has been found to be appropriate. Temperatures below 18°C may induce a cold stress response and shivering, resulting in 'flicker' thermograms. While temperatures above 20°C may cause sweating and create other anomalies and noise on the image. It is also important for the area of the body under examination to reach a steady state in controlled environmental conditions. In some cases, it is necessary for the patient to partially undress so as to expose the area of the body to be examined directly to the ambient temperature for a short stabilisation period (10 -15 minutes is usually enough). Loose clothing and underwear are required to avoid altering the local blood flow (and thus the overlying skin temperature) through the pressure and restriction caused by tight-fitting garments. Dressings, ointment or any other surface moisture will accentuate, to a certain degree, the infra-red emission from the skin and must be eliminated prior to equilibration if thermograms are to be used in a controlled manner. This sensitivity is due to the fact that infra-red radiation in the wavelength typically used (between i.e. 3-5 and 8-12 $\mu\text{m}$ ) are strongly absorbed by water and organic compounds [32].

If environmental conditions are adequately controlled, heat emission from the skin is largely determined by the underlying blood supply. In the absence of deeper lying organic disease or other factors, which may indirectly alter skin blood flow, the thermographic image of the heat emitted by the skin may be interpreted in terms of the status of the underlying peripheral circulation [33-36]. Thermography can therefore be used for detecting peripheral arterial disease, diabetic angiopathy, Raynaud's phenomenon and related conditions, inflammatory conditions, and determination of amputation level etc. For deeper-seated pathological conditions, radio-isotope imaging, ultrasound or radiography is more suitable.

At present IR thermography is most widely used in applications associated with the vascular system [37,38], peripheral and cutaneous circulations as well as relatively superficial tissue lesions. In some cases thermography provides a beneficial preliminary or complementary aid to examination, in others it fills in gaps in the existing armoury of assessments. However, its use in clinical



assessment is still considered by some to be controversial, partly due to the wide range of temperatures of lesions or diseases (this is particularly true in breasts) and also due to the lack of understanding of the basic principles of thermography (i.e. characteristics and limitations). The work described in sections 7 addresses some of these recognised problems.

### ***5.2 Static and Temporal Sequences of Thermograms***

Various methods have been used for the presentation, analysis and classification of thermograms, these include; functional images [33], spatial signatures and texture signatures. Historically these methods have been applied to individual static thermograms for diagnosis. Many conditions, however, are not evident from such single static images. Historically, also, the assessments from a static thermogram would normally be based on an individual patient's data only, much like the common use of X-ray pictures. Diagnostic results would be in the form of an index or some form of written report, with or without graphical explanation. These approaches do not fully utilise the information available from thermography.

It has been found that useful information can be obtained by observing the thermal behaviour of the body over time. In order to do this the technique of temperature stress testing has been developed whereby a temporal thermal response is induced in the body under controlled conditions. A sequence of thermograms is taken to record the body's thermal behaviour and the correspondence between different thermograms is built up by aligning the body in the images with a stretching and aligning process. An elastic stretching approach described in the following section was developed to address the alignment problem described. Two diagnostic systems will be described both of which use analysis of the temporal thermal response and wider statistical data for automatic classification and diagnosis. The first, in section 6, is concerned with diagnosing joint disease, namely arthritis, Raynaud's syndrome and rheumatism. It is based on analysing the body's thermal behaviour after a 'cold stress'. The second system, described in section 7, is concerned with the diagnosis of breast cancer after a 'heat stress'.

### ***5.3 Advantages and Disadvantages of Thermography in Medical Applications***

There are numerous advantages in applying infra-red thermography in clinical and medical investigations. They include:

- The recording of a patient's thermogram is inexpensive to perform, although initial equipment costs can be relatively high.
- The technique is simple, and can be repeated at frequent intervals, allowing real time assessment (especially compared with radiological images), this often results in good patient co-operation.
- It is non-invasive and involves no radiation hazard.
- It is a non-contact imaging approach
- Thermography may indicate lesion too small to be seen on a roentgenogram.

- Each patient has a unique thermogram, useful in a follow-up analysis.
- Thermographic images are opaque and do not contain the overlapping objects present in radiographic images.
- Thermographic images are inherently quantitative in nature due to the direct representation of the physical temperature.
- Only rudimentary knowledge of anatomy and pathology is necessary to interpret a thermogram. Diagnostic criteria involve essentially measures of symmetry and relative temperature distributions and patterns.
- Since the temperature signal is available directly in an electrical form, it is simple to connect the thermographic equipment directly to a computer via an analogue-to-digital converter for automatic image acquisition and storage.

There are, of course, disadvantages, which include:

- Lack of specificity.
- Provides limited range of image brightness.
- Low spatial resolution.
- Not all abnormalities exhibit observable thermal phenomena.
- Thermal variations do not necessarily have a spatial relationship to the disease investigated.
- Thermographic signs can occur in benign as well as malignant conditions (e.g. breast thermography).
- Occasionally, an anatomical aberration can give a false positive reading.
- Simultaneous, bi-lateral symptoms could be diagnosed as negative.

## **6. Joint Disease Diagnosis Using Dynamic Infra-red Thermography**

### ***6.1 System Overview***

The objective of this work was to automate the analysis and classification of a temporal series of thermograms recording the response of the hand to a cold stress into different classes, namely, normal, Raynaud's and inflammatory. Various other quantitative imaging systems have been used for assessment of inflammation in joint disease, for example: the differential thermistor thermometer, the infra-red thermometer, radiography and arteriography. Their limitations lie in the technical difficulties, expenses and/or their invasive nature.

In this system statistical pattern recognition techniques are used to analyse the results of a temperature stress test. Immersing the hand in a cold water bath at 20°C for 1 minute (cold stress) induces the temperature stress. The body's response to the stress is recorded by thermograms taken at regular two-minute intervals over about twenty minutes. In order to study the thermal response of the hand it is, in general, necessary to track the thermal behaviour over time of every point on the hand. It is therefore necessary to have knowledge of the correspondence between points in each of the series of images. The simplest way to do this would be to ensure that there was no movement of the hand over the twenty minutes period, thus the thermal response of each object pixel could be tracked in a straightforward manner. Unfortunately, it is not possible to restrain the hand in a way that does not affect its thermal response. The patient must

simply hold the hand in front of the camera and the correspondence must be built up by aligning the images of the hand with a stretching and aligning algorithm [39].

The thermograms used in this study were taken at a distance between 0.75 – 1.5 metres, depending on the size of hand in question. The thermograms were digitised into a  $128 \times 128$  square image with 256 grey-levels. In this work only the anterior view of the hand (back of the hand) was used, and no attempt was made to analyse thermograms taken from posterior or lateral views.

The feature extraction method developed in this application is based on the Kittler & Young transformation [33]. The function of which is to extract those features that appear most important for classification. A 7-Nearest Neighbour (7-NN) classifier, built using the Condensed Nearest Neighbour (CNN) technique, is applied to the features. It was recognised that it would be desirable for the resultant classification (*diagnosis*) to be presented as a colour coded diagnostic image not only to be classified into appropriate disease categories, but also for the ‘hot spot’ (inflammatory condition), or ‘cold spot’ (Raynaud’s) to be identified appropriately. The severity of the disease should be indicated through the intensity of the colour in the diagnostic image on a display. For example, using green colour for normal case, red for inflammatory cases and blue for Raynaud’s. In the case of uncertainty, the resultant image would have a non-primary colour. The system achieved about 96% accuracy at pixel level.

### ***6.1 Iterative Elastic Object Stretching and Aligning***

The objective of this technique is to align the series of thermograms of hands [39] for subsequent analysis. The series of thermograms are taken over a period of approximately 20 minutes in order to study the thermal response of the body to a temperature stress. The temperature stress is induced by immersing the hand in a cold water bath at 20°C for 1 minute (*cold stress*). The body's response to the stress is recorded by thermograms taken at regular two-minute intervals over about twenty minutes. In order to study the thermal response of the hand it is, in general, necessary to track the thermal behaviour over time of every point on the hand. It is therefore necessary to have knowledge of the correspondence between points in each of the series of images. The simplest way to do this would be to ensure that there was no movement of the hand over the twenty minutes period, thus the thermal response of each object pixel could be tracked in a straightforward manner. Unfortunately, it is not possible to restrain the hand in a way that does not affect its thermal response. The patient must simply hold the hand in front of the camera and the correspondence must be built up by aligning the images of the hand with a stretching and aligning algorithm.

The thermograms used in this study were taken at a distance between 0.75 – 1.5 metres, depending on the size of hand in question. The thermograms were digitised into a  $128 \times 128$  square image with 256 grey-levels. In this work only the anterior view of the hand (back of the hand) was used, and no attempt was made to analyse thermograms taken from posterior or lateral views, Figure 7.

The approach developed was based on the ideas first presented by [40] and based on the basic principle of Burr's technique [41, 42]. It works on a pair of hand image, one rigid, and the other elastic image. The elastic image is iteratively stretched and relaxed by "forces" derived from local feature association between the elastic image and a rigid target image. Two sets of "Local Difference Vector" (LDV) are derived using nearest neighbour relationships between the outline pixels of the two hand images. One set from the elastic image to the target image and the other set from the target to the elastic image (*the magnitudes of these vectors decrease as misregistration decreases*). Weighted sums of local LDVs are used to create "Corrected Difference Vectors" (CDV). These CDVs are used as the "forces" applied to the whole the elastic image to create an intermediate image for further matching with the target images.

The elastic image is iteratively distorted to match the target at each iteration, the rigidity of the elastic image is reduced giving a sequence of modified images, each a better match to the target. At the beginning the image pair can be expected to be vastly mis-aligned and therefore initially rigid matching is preferred so as to avoid incorrect alignment. However, as the images become more closely matched finer details and more localised feature alignment can be achieved with increased reliability.

Although there is no theoretical proof of convergence, experiments show that mis-registration decreases with each iteration.

### 6.1.1 Outline Derivation

Derivation of the outline is necessary for two reasons. Firstly the temperature pattern on the hand varies with time, it is therefore, impossible to attempt to align the series of images on the basis of the whole hand, where there is nothing to align. The outline of the hand must therefore be used to provide accurate and representative information about the position of the hand, so as to allow matching to be performed efficiently. Secondly elastic matching is computationally intensive and so it is desirable to minimise the number of pixels, which participated in the matching process.

Here, a single pixel width outline is derive using an adjacent neighbour thresholding technique which is simple and provides a single pixel wide outline suitable for the stretching process.

The input image is converted into a binary image using a straightforward histogram based threshold as it has a very uniform background. A value "1" is assigned to all object pixels of the hand (*displayed in white*), while a "0" is assigned to all background pixels (*displayed in black*) as shown in Figure 9a. The sum of the four adjacent neighbours of each hand pixels (above, below, left and right) is calculated in Figure 9b. If the sum is 4 then the corresponding pixel is labelled as the background and is assigned a value of "1", otherwise "0" is assigned for the outline of the hand. At the end of this process, an image of a single pixel wide hand outline is produced, with the pixel coordinates recorded

for the matching process. The base line of the hand was not included in the matching, as shown in Figure 9c, to further reduce the dimension of the outline.

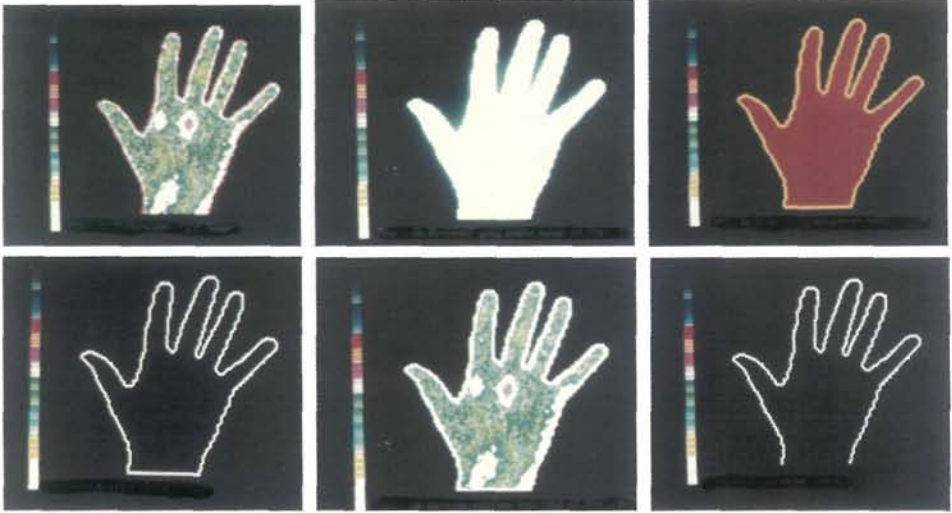


Fig. 9a - 9f

### 6.1.2 Dynamic Corrective Matching

#### 6.1.2.1 Local Difference Vectors

Define the outline of the hand in Image  $I_1$  (i.e. the elastic image) as the string:

$$OL_1 = IX_1(n), IY_1(n), \text{ where } n=1, \dots, N \quad (20)$$

$IX_1(n), IY_1(n)$  are the X, Y coordinates of the  $n^{\text{th}}$  outline pixel, and N is the total number of outline pixels in  $OL_1$  is defined as:

$$OL_2 = IX_2(m), IY_2(m), \text{ where } m=1, \dots, M \quad (21)$$

M is the total number of outline pixels in  $OL_2$ . The distanced DNM between the  $n^{\text{th}}$  element of  $OL_1$  and the  $m^{\text{th}}$  element of  $OL_2$  is defined as:

$$DNM = |IX_1(n) - IX_2(m)| + |IY_1(n) - IY_2(m)| \quad (22)$$

For a given outline pixel, n, the corresponding point  $m'$  on the other outline for which DNM is minimum is the nearest neighbour giving the local difference vector  $LDV_1$  at point n in image  $I_1$ :

$$LDV_1X(n) = (IX_2(m') - IX_1(n)) \text{ in x direction} \quad (23)$$

$$LDV_1Y(n) = (IY_2(m') - IY_1(n)) \text{ in y direction} \quad (24)$$

Similarly, for image  $I_2$ , the  $LDV_2$  for the x and y directions are:

$$LDV_2X(m) = (IX_1(n') - IX_2(m)) \text{ in x direction} \quad (25)$$

$$LDV_2Y(m) = (IY_1(n') - IY_2(m)) \text{ in y direction} \quad (26)$$

Where  $n'$  is defined as the index  $n$  ( $n=1, \dots, N$ ) that minimised DNM for a fixed  $m$  in  $I_2$ . The  $x$  and  $y$  local difference vectors are treated independently so as to ensure that  $x$  and  $y$  movements are not dependent on each other.

In general the vectors  $LDV_1$  will be different from the vectors  $LDV_2$  (Figure 10) and the local errors in the LDVs could lead to extreme mismatches especially around sharp corners, i.e. the nearest neighbour pair for  $I_1$  with respect to  $I_2$  is not necessarily the same as the corresponding nearest neighbour pair for  $I_2$  with respect to  $I_1$ . Matching association using such error prone or misleading difference is unreliable and disappointing. In order to minimise such errors, both  $LDV_1$  and  $LDV_2$  are used to compute a corrected difference estimate where the two measures complement each other, the combination of both helps to pull together the mis-alignments.

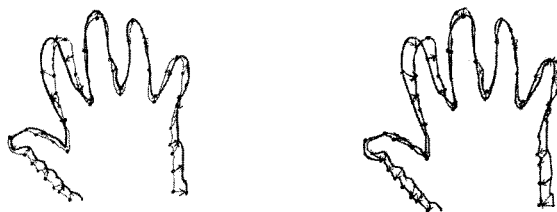


Fig. 10:  $LDV_1$  of  $I_1$  and  $LDV_2$  of  $I_2$  (Black  $I_1$  and Red  $I_2$ )

### 6.1.2 Corrective Difference Vectors

In order to achieve a desirable match the errors found between the corresponding local nearest neighbour difference vectors (LDVs) have to be eliminated or at least minimised. Furthermore, if the LDVs were to be a complete set of labels for relaxation updating it can easily be seen that it is necessary to reduce the computational load and storage requirements. Both these requirements can be met by realising that in general objects such as hand pixels lying close to an outline should be more influenced by its movement than by movement of outline pixels further away. To effect this an averaging scheme can be used in which a Gaussian weighting with distance allows a gradual decrease in effect with distance. Thus a corrected difference vector (CDV) at an image point  $(X, Y)$  in  $I_1$  can be defined in  $x$  and  $y$  directions, i.e.  $CDV_x$  and  $CDV_y$ :

$$CDV_x(X, Y) = (\sum_{n=1}^N F_1(n) * LDV_1X(n)) / (2 \sum_{n=1}^N F_1(n)) - (\sum_{m=1}^M F_2(m) * LDV_2X(m)) / (2 \sum_{m=1}^M F_2(m)) \quad (27)$$

$$CDV_y(X, Y) = (\sum_{n=1}^N F_1(n) * LDV_1Y(n)) / (2 \sum_{n=1}^N F_1(n)) - (\sum_{m=1}^M F_2(m) * LDV_2Y(m)) / (2 \sum_{m=1}^M F_2(m)) \quad (28)$$

Where  $N$  and  $M$  are the number of outline pixels in  $I_1$  and  $I_2$  respectively, and  $F_1(n)$  and  $F_2(m)$  are the Gaussian weighting functions define as:

$$F_1(n) = \exp - (X - IX_1(n))^2 - (Y - IY_1(n))^2 / SCL^2_i \quad (29)$$

$$F_2(m) = \exp - (X - IX_2(m))^2 - (Y - IY_2(m))^2 / SCL^2_{I_2} \quad (30)$$

LDV<sub>1</sub>(n) and LDV<sub>s</sub>(m) are the local difference vectors at points n and m in I<sub>1</sub> and I<sub>2</sub> respectively. The coordinates (IX<sub>1</sub>(n), IY<sub>1</sub>(n)) and (IX<sub>2</sub>(m), IY<sub>2</sub>(m)) represent the location of the outline pixels in n and m respectively in the two images and SCL is the scaling factor.

The effective correcting SCL, should initially be roughly the dimension of the image to achieve an overall correction of mean difference between the two image, i.e. a rigid translation. This is because, as mentioned above, detailed matching at an early stage can lead to erroneous registration due to incorrect alignments. Subsequently SCL should be decreased at each iteration. The average difference between the corrected difference vectors varies inversely with SCL. The SCL used in this application is determined dynamically based on the amount of difference between the two image pair and hence SCL for the iteration is defined:

$$SCL_i = \sqrt{\sum_{n=1}^N (LDV_1X(n)^2 + LDV_1Y(n)^2)} \quad (31)$$

The CDV of I<sub>1</sub> is evaluated for all the hand pixels, taking into account the iteration between the target outline (OL<sub>2</sub>) and its outline (OL<sub>1</sub>). It is then used to calculate the new stretched coordinated position for each corresponding pixels. The new stretched coordinate position in image I<sub>1</sub> is defined as:

$$S(X,Y) = I_1(X,Y) + CDV(X,Y) \quad (32)$$

Which can then be used for further matching with respect to I<sub>2</sub>.

The iterative process terminates either when a pre-defined satisfactory match is achieved or if the stretching process begins to oscillate. The latter situation might arise when the object is half way between the target outline (Figure 11).

By restricting the local difference vectors, LDVs, to the nearest neighbours the influence exerted by other local vectors is ignored. Since deformation is restrictive in this application, i.e. the possible different positionings and movements of the fingers and the rest of the hand are limited, a straightforward nearest neighbour measure proved to be appropriate. It is possible that for more strongly deformed and complex images, more sophisticated local feature measure would be required, e.g. rotation. The averaging method, equations 27 and 28, was chosen because mis-matches counteract each other on average while correct matches enhance each other, since they are well correlated.

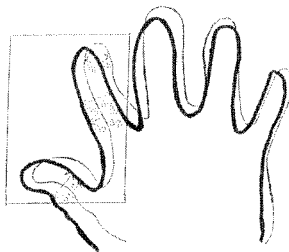


Fig. 11: Oscillation occurs when matching between two equidistant target outlines

The summation equations 27 and 28 can be calculated over all the object pixels in each image or restricted within a neighbourhood, for example, a  $30 \times 30$  pixel. However, using the square meant that the area of influence was uneven, varying from a maximum, the diagonal to the four corners, to a minimum, the perpendicular to the centre of the four sides. In order to produce a more evenly distributed area of influence and increase the area of influence while also reducing the computational load, a more even approach was developed. In this even approach a circular Gaussian function with the radius of half the width of the image area, i.e. covering the entire image, was used. Smaller radii were tried to limit the computation but were found to be unsatisfactory.

The two-dimensional Gaussian weighting function  $G(X,Y)$  in this application is dynamic, i.e. it is different from each iteration to iteration and dependent on the degree of alignment between the two images, so that more detailed or localised matching can be achieved by adjusting the scale value,  $\sigma$  in  $G(X,Y)$

$$G_i(X,Y) = 1/\sqrt{2\pi} \exp - ((x^2 - R)^2 + (y^2 - R)^2)/2\sigma^2 \quad (34)$$

Centred at  $(R,R)$  and where:

$$\sigma = \text{FIT}(I)/2.57 \quad (35)$$

FIT as defined in equation is associated with the 99% confidence level.

This Gaussian distribution can be viewed as an image (Figure 12) as can the CDVs. Consider the  $\text{CDV}(X_i, Y_i)$  calculated at the  $j^{\text{th}}$  outline pixel, i.e. at  $(X_i, Y_i)$  the displacement in the  $x$  and  $y$  directions at a pixel given separately are independent of each other. In fact, these  $x$  and  $y$  directional displacements (CDVs) at the outline's pixels can be viewed as two separate  $x$  and  $y$  images, as shown in Figure 13. The value at each outline pixel becomes the amount of appropriate directional displacement at that pixel. Thus, there are 3 images available: the movements at the outline in the  $x$  and  $y$  directions, and the Gaussian weighting function. The movement of the interior of the hand is determined by the movement of its outline. In effect, each hand pixel's movement is a linear combination of all the outline pixels' movements scaled by the Gaussian weighting function as defined:

$$\text{CDV}'_x(X,Y) = \sum_{\beta=1}^N \sum_{\alpha=1}^N \text{CDV}_x(\alpha,\beta)(X-\alpha, Y-\beta) \text{ in } x \text{ direction} \quad (36)$$

$$\text{CDV}'_y(X,Y) = \sum_{\beta=1}^N \sum_{\alpha=1}^N \text{CDV}_y(\alpha,\beta)(X-\alpha, Y-\beta) \text{ in } y \text{ direction} \quad (37)$$

Equations 32 and 33 represent the convolution of the two-dimensional Gaussian function with the CDVs in  $x$  and  $y$  directions respectively. Such spatial convolutions are computationally expensive. The computationally cheaper frequency domain technique of convolution was therefore applied in this application to evaluate  $\text{CDV}'_x(X,Y)$  and  $\text{CDV}'_y(X,Y)$ .



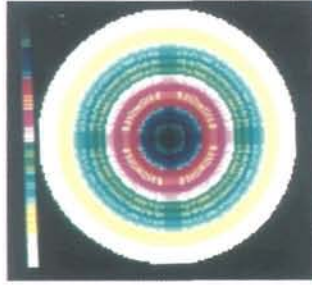


Fig. 12: 2-Dimensional Gaussian Weighting Function

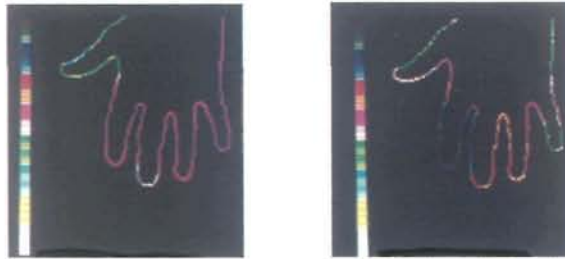


Fig. 13: CDVs evaluated at outline in x Direction and y Direction

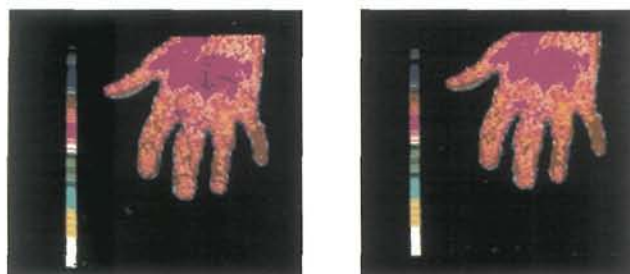
Briefly, the required convolution is achieved by taking the two-dimensional Fourier transformation of each CDV image (in x and y directions) and  $G(X,Y)$ , multiplying the transformations together, and then calculating the inverse transformation of the result.

The choice of two independent CDV images rather than for instance, using the complex notation (real for x direction and imagery for y directions), means that x and y movements are independent. This is an important feature of the stretching approach, which allows the desired stretching to be obtained by means of the combined effect of the independent x and y movements. The resultant  $CDV'_x$  and  $CDV'_y$  are then used to stretch the elastic image.

#### 6.1.4 Stretched Image Refinement Process

When the elastic image is aligned iteratively with a target image through intermediate "stretched" images some of the pixels in the stretched image might be left out due to the compression, thus leaving gaps in the hands, Figure 14a, while some new pixels might be generated for expansion. These new expanded pixels could in general be stretched outside the boundary of the image, and should be ignored. It is however, necessary to replace "blank" pixels in the hand. The blank pixel value is replaced by the average of its 3x3 hand pixel values if the total number of the non-blank hand pixels is more than three, otherwise it is left (i.e. becomes part of the background). This is deemed a logical approach as all the pixels (temperatures) in the hand are related to each other and influenced by each other. This refinement process was applied to all the examples in this

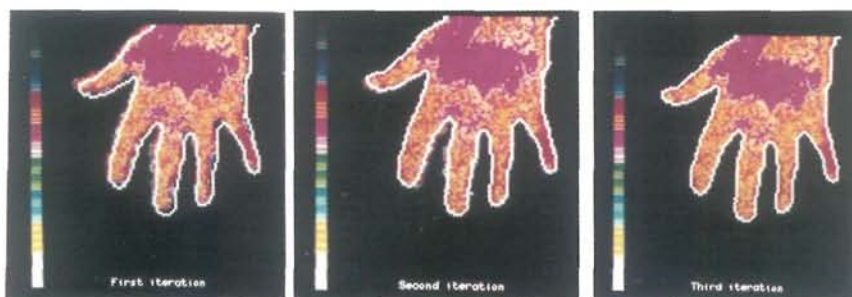
sections and the refined version of the image in Figure 14a is shown in Figure 14b.



*Fig. 14a and 14b: Refinement Scheme*

### 6.1.5 Experimental Results and Discussion

As can be observed in Figure 15, the detailed matching of the frequency domain elastic matching approach perform well in matching a pair of hands taken at different times. In general it takes only 3 iterations to achieve a good match. This dynamic approach has been shown to be an effective means of automatic matching of two or more objects based on gradual and sequential deformation of an elastic image onto a target image. An efficient but simple nearest neighbour rule is used to supply sufficient information to carry out the matching.



*Fig. 15: Matching Images*

## 6.2 Thermographic Colour Coding Schemes

### 6.2.1 Gradual Natural Scale (GN)

The use of grey-scale (e.g. Figure 16) or pseudo-colour (e.g. Figure 17) in medical images has been the subject of debate [33]. Some believe that colour-coding is artificial and mis-leading and can create confusion which leads to mis-interpretation. While others find that the use of grey-scale in some images makes it difficult to differentiate pathological areas from normal areas. This controversy could partly be due to the use of inappropriate pseudo-colouring systems which are insensitive to the particular information required and the requirement for special expertise to interpret the colour codes.



*Fig. 16: A Grey-Scale Hand Thermogram*

Two different colour coding schemes are considered here [40]. The first, Gradual Natural (GN) coding scheme is shown in figure 17. It is a smooth gradual scale ranging from black (cold) to white (hot) with blue, green and red gradually intermixing with one another over the inter-mediate range. Such a scheme conveys the overall temperature range of an image in an easily identified and recognised colour spectrum, giving a general idea of where the temperatures lie. The semi-circular disc shown on the left of the image is the temperature standard.

The encoding is based on three primary colours - red, green and blue, and has 8 bits to represent all the colours, i.e. 256 levels. These levels are split into 3 ranges, one for each colour, and within the range the intensity can be varied uniformly and gradually. At the boundary between any two primary colours (e.g. green and red), a gradual mixing of the two colours (e.g. a decrease in green colour intensity accompanied by an increase in red colour intensity) results in the perception of a non-primary colour (e.g. yellow). This non-primary colour is necessary to create smooth changes over the boundary of the two different colours, this providing an overall gradual and smooth colour spectrum.

### 6.2.2 Randomised Blocking Scale (RB)

The second coding scheme, Randomised Blocking, is illustrated in Figure 18; as its name suggests it uses randomised colour blocks. As before black and white are at either end but the intermediate colours are small repetitive blocks with different colours and intensity. Adjacent temperatures are represented by significantly different colours, so that slight temperature differences will be accentuated; such differences would otherwise be undetected.

The coding scheme is constructed as follows. The 8-bit control byte is split into three fields, one for each primary colour; red, green and blue. Thus in the example of Table 1, bits 0, 3, and 6 are associated with red (R), bits 1, 4, and 7 for green (G) and bits 2 and 5 for blue (B). This 3-3-2 combination means that there are 7 possible intensities for both red and green but only 3 for blue. An example is given in Table 1, here level 7 of green is combined with level 7 of red and no blue, e.g. resulting in a yellow colour.



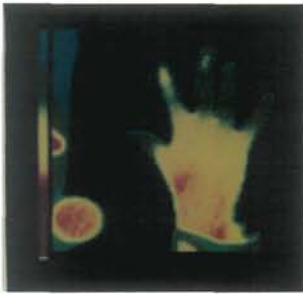


Fig. 17: Gradual Natural Scale

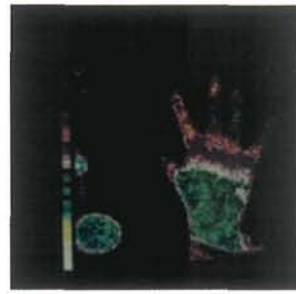


Fig. 18: Randomised Blocking Scale

<i>Bit</i>	7	6	5	4	3	2	1	0
<b>Colour</b>	G	R	B	G	R	B	G	R
<b>Example</b>	1	1	0	1	1	0	1	1

Table 1

This coding scheme is also useful for coding disease classification categories because in such applications there is not necessarily a uniform continuum of information to be encoded (i.e. just blue of some degree, just red of some degree or just green) but rather, as in this case, any possible combination of red and green or blue and green.

### 6.3 Observed Hand Thermal Characteristics

During the course of data collection, some thermal characteristics or behaviour patterns were noted. It is these patterns that the system must be able to extract, analyse and quantify.

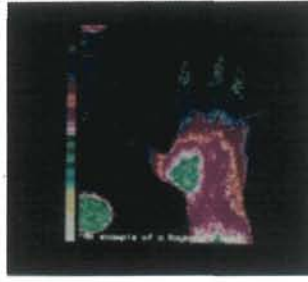
#### 6.3.1 Normal Condition

It is known that in a normal hand all the fingers remain at similar temperatures ( $27^{\circ}\text{C} \pm 2^{\circ}\text{C}$ ) and display a positive gradient, i.e. temperature increases towards the finger tips. Most showed hypothermia during the 15 – 20 minutes stabilisation period, which resulted in an increase in hand temperature and which decreased the temperature differences between a normal and an inflamed hand.

After the cold stress the hand rewarmed quickly and throughout the test the hand maintained its temperature distribution with hotter fingers. An example of a normal hand is shown in Figure 19. In some normal cases a diagonal gradient pattern could be noted while in other cases isolated cold fingers were found which did not necessarily relate to the symptoms.



*Fig. 19: Normal Hand*



*Fig. 20: Raynaud's Hand*



*Fig. 21: Inflammatory Hand*

### 6.3.2 Raynaud's Conditions

In Raynaud's phenomenon (primary and secondary) the mean temperature was always lower than  $27^{\circ}\text{C}$  (approximately  $22^{\circ}\text{C} - 23^{\circ}\text{C}$ ) and during the cold stress the temperature range across the hand was up to  $10^{\circ}\text{C}$ , whereas in normal hands it was no more than  $6^{\circ}\text{C}$ . The symptoms were characterised by a well banded isothermal pattern with negative gradient (i.e. colder towards the finger tips). In some patients with severe cases of secondary Raynauds' the gradient could be over as much as  $12^{\circ}\text{C}$ . Similar, but less marked, banded patterns over the hand were found in patients with primary Raynaud's. The hands of patients with Raynaud's condition tended to cool down during the stabilisation period prior to the cold stress. In patients with primary Raynaud's, the hands rewarmed after the stress and in the end the gradient differences were reduced. In patients with severe secondary Raynaud's they could cool down further in response to the cold stress. An example is given in Figure 20.

### 6.3.3 Inflammatory Conditions

Higher temperature is often recorded ( $29^{\circ}\text{C} - 34^{\circ}\text{C}$ ) for arthritic hands due to the inflammatory mechanism. Classic symptoms being swelling, slight deformity and the presence of a "hot spot". The temperature rise on the overlying skin at the affected joint can be up to  $5^{\circ}\text{C}$ . The precise nature and extent of such hypothermic areas are determined by the underlying pathology. For example, synovitis may cause a localised area of increased temperature, while chronic rheumatoid arthritis may result in a generalised hypothermia over the whole joint. Gout also causes dramatic and characteristic increase in temperature over affected joints. Inflamed areas remained at higher temperatures after cold stress. During the "warm up" period the affected hand warmed up as in the normal hand but less markedly. In fact, due to vasodilation, the temperature difference between the two classes was less prominent in the early stage of cold stress response [38]. An example is given in Figure 21.

## **6.4 Application of the Kittler & Young Method to the Analysis of Sequences of Hand Thermograms**

### **6.4.1 The Kittler & Young Method**

Ideal classification criterion should be based on the discriminatory potential of both the class means and class variances, scaling them according to their relative significance. Such criteria are, however, complex and difficult to formulate. In many practical situations the contribution of the differential ability of class variances is neglected. This simplicity is assumed by all standard variants of the Karhunen-Loeve transformation based methods [43]. The Kittler & Young method [44] is a feature extraction method based on two Karhunen-Loeve transformations, and is intended to overcome this problem.

### **6.4.2 The Application of Kittler & Young Method**

The Kittler & Young method was applied to a series of thermograms (e.g. those in Figure 22) to compress the differential thermal information, based on both the thermal means and variances (of different classes), of thermograms into the first few transformed "Eigen" images.

In order to accentuate the features of the first 3 resultant transformed images for visual inspection and for further use as diagnostic images, the following colour coding scheme, based on the RB scale, was applied:

1. The first transformed image (Figure 23) was coded into 7 different levels of green, where the maximum value corresponded to the darkest green and vice versa. The 7 levels of green are given the following values: 0, 2, 16, 18, 128, 130, 144, 146 respectively. These values coincided with the colour 'bits' assignment in Table 1.
2. The second transformed image (Figure 24) was coded into 7 different levels of red, as described above but given the following values: 0, 1, 8, 9, 64, 65, 72, 73.
3. The third transformed image (Figure 25) was coded into 3 different levels of blue, 0, 4, 32, 36.

Individually the three colour coded transformed images with different intensities of the corresponding colour accentuates the variations in thermal behaviour of different parts of the hand, with the highest intensity corresponding to the maximum variation. Among them the green image conveyed the most discriminatory information about the variation in the corresponding parts of the hand over the time series, as well as within the hand anatomy.

The coding matched quite well with the diagnosis of the physician and could be used as a general guidance diagnostic image. In order to make the most of the three transformed images, however, a composite transformed diagnostic image was developed (Figure 26). This is created by summing the three colour-coded eigen-images together. From this composite image, the difference in thermal responses in the hand could be seen more clearly than in any of the three colour coded transformed images individually. The "hot spot" (in inflammatory conditions) and "cold spot" (in Raynaud's conditions) could be identified more

easily than using the three individual transformed images. This composite transformed image was found to match closely the clinician's diagnosis.



*Fig. 22: Series of Thermograms*



*Fig. 23: First Transformed Image Coded in Green*



*Fig. 24: Second Transformed Image Coded in Red*



*Fig. 25: Third Transformed Image Coded in Blue*



*Fig. 26: Composite Transformed Image*

## 6.5 Classification

### 6.5.1 Training and Testing Data

The available data came from three diagnostic classes; inflammatory, normal and Raynaud's. Some of the affected areas were localised, though mixed classes commonly occur in the same patient (for example only parts of the hand might be affected by a disease while the rest of the hand is normal). The training data for the Kittler & Young therefore consisted of appropriate diagnostic classes only, not all the pixels in a classified hand. This meant, for example, that only the inflamed areas of a hand were used in the inflammatory training set, and similarly, for the normal and Raynaud's classes. The training data thus contained only representative data of its class and was not mixed with other classes. The selection of different class representative vectors was carried out by visual inspection of the diagnostic composite eigen-images. Square or rectangular areas were annotated manually on the displayed composite eigen-image and the corresponding regions in the original series of thermograms were then extracted. These selected areas formed the training data and the optimal co-ordinate system was then obtained by applying the Kittler & Young analysis to this data.

In order to reduce the storage and computational requirement for the classification a Condensed Nearest Neighbour (CNN) classifier [45] was used.

### 6.5.2 Classification Results

Three CNN classifiers were built based respectively on the Eigen images corresponding to the first 2, 3, and 4 Eigen vectors. These were tested using the 7-NN classification technique with a Euclidean distance measure and a majority vote system. The testing samples were transformed using the eigen-vector matrix derived from the training data set. It was found that the best performance was achieved when three-dimensional CNN classifier was used giving an error rate of only 4.3% in terms of pixels as classified by a clinician. It was found that the majority of errors came from the mis-classification of the Raynaud's class as normal. This is possibly due to the similarity in behaviour of a "cold had complaint" and that of a mild Raynaud's condition.

It can be concluded from the above that only 3 dimensions are needed because the Kittler & Young method has succeeded in compressing most of the discriminatory information into the first three components of the transformed feature vectors, making the rest of the components in the transformed space redundant.

The performance of the system compares favourably with a clinician's diagnosis. Moreover, the majority of the errors came from the 4-3 tie condition of the 7-NN classifier where the correct category was in the 3-minor vote position. Hence, although they were mis-classified their correct classification could still be identified when presented in the colour representation technique developed in this study and described below.



## 6.6 Presentation of Classification Results for Diagnosis

The resultant classified thermal features were used to produce diagnostic images; in these the colour coded "classification pixels" or "diagnostic pixels" were used to replace the corresponding pixels in a thermogram. The degree of confidence in the classification was denoted by the different intensity of the assigned disease class colour. For example, in the 4-3 tie condition of normal-Raynaud's classification, then 4th degree intensity of green (normal) together with 3rd degree intensity of blue (Raynaud's) were summed together, giving a 'greenish - blue' colour as the final 'diagnosis colour' to the pixel in question. In fact this 'greenish-blue' colour was found to be associated closely to those patients with complaints of cold hands (so to some extent a mild primary Raynaud's). While 'yellowish' cases were found to be the patients with mild inflammatory conditions, i.e. when there was a tie between normal (green) and inflammatory (red).

The resultant diagnostic images in Figure 27 indicated the locations of the affected areas on the hands as well as the degree of "truthfulness" (*severity and certainty*) of classification by means of different colours and varying degrees of intensity.

The diagnostic results compared extremely well with the clinician's diagnosis. The effect of the error rate was that the exact dimensions of the affected area might not be precisely defined, but the locations were identified. The classification errors were most likely to be at the boundary between two different classes on the hand. The majority of errors came from assigning Raynaud's class into normal control class, i.e. the group of people with cold hand complaint.

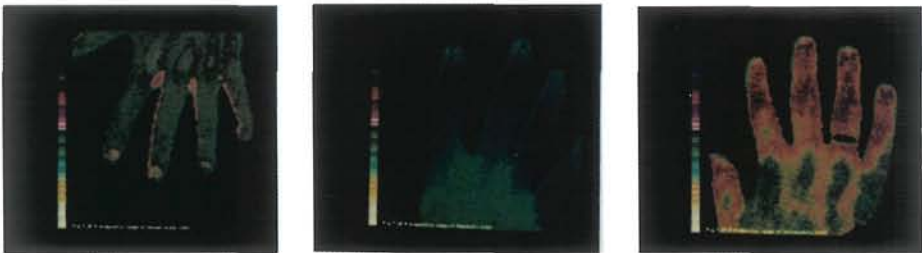


Fig. 27: Classified Normal Hand, Raynaud's Hand and Inflammatory Hand

## 7. Breast Cancer Detection

### 7.1 Breast Cancer Detection

Every year 1500 women in the UK alone die of breast cancer [46, 47]. The success of a breast screening programme depends heavily on the quality/effectiveness of the clinical service. At present, when the assessment is carried out each speciality (radiography, surgery, pathology and ultrasonography) carries out its examination/test independently. Studies have shown that early

detection is crucial to survival in breast cancer, and it is believed that the use of appropriate image processing can make screening and diagnosis easier. Currently X-ray mammography is the most commonly used imaging technique for diagnosis [45]. Due to their invasive nature X-ray mammograms can not be acquired regularly. An alternative complementary technique using non-invasive infra-red thermography [48] together with low-level microwave radiation is described here [50]. Neither thermography nor mammography can detect breast cancer, they are used for indicating abnormalities which could be related to cancer or other diseases, such as fibrocystic disease, infection, etc.

Thermography, unlike x-ray mammography, does not involve painful breast compression, i.e. it is a non-contact imaging procedure. Furthermore, it is non-invasive, it does not use any radiation, and can be repeated as often as required.

There are many means of collecting breast thermograms, but fundamentally the breast image is captured by an infra-red detector together with an infra-red imaging system, typically with an 8-bit dynamic range, a 100mK sensitivity at 30°C and a spectral range of 8-12 microns. The infra-red thermogram can then be examined by a clinician [48].

Typically, [48], the breast thermogram is divided into four quadrants, namely; upper outer, upper inner, lower outer and lower inner by drawing lines on the infra-red images from the chin of the patient to each nipple and then two horizontal lines left and right to the edge of the breast. Finally, a fourth line is drawn to the lowest contour of the breast. The mean, standard deviation, minimum, and maximum temperature are measured for each quadrant of both breasts and then they are compared to detect any differences between breasts. In [48] it was shown that if the mean temperature difference is equal to or greater than 0.5 °C, then the patient is considered to exhibit an "abnormal" asymmetrical thermal pattern. It was found that in the cases examined breast abnormalities were detected before they could be detected by self-examination or by mammograms. However, this approach poses a problem when both breasts are abnormal and exhibit symmetrical thermal patterns, then the underlying problems will remain undetected until they manifest themselves through self-examination or in the mammograms.

## ***7.2 System Overview***

This work is concerned with the thermal analysis of normal and abnormal infra-red mammograms [50]. The analysis is based on a series of infra-red mammograms of the breast subjected to a warming temperature stress. The temperature stress is induced using a very low level of microwave radiation. This produces a higher heating rate in tumourous tissue than in normal healthy tissue. The technique uses a sequence of mammograms taken during the cooling process to classify abnormal conditions that manifest themselves as body temperature abnormalities. The image analysis and processing is basically the same as that described for the diagnosis of joint diseases in Section 6 above.

### 7.3 Temperature Stress Technique and Data Collection

It has been found that the static infra-red mammogram provides insufficient information for diagnosis due to the limited transmission of IR radiation through fatty tissues. Therefore, in this study again a temperature stress is applied to induce a thermal response of the breasts to aid diagnosis. Here the body is exposed to a micro-wave heating (frequency: 0.45Ghz, power density: 80 -100m W/cm<sup>2</sup>) for two minutes. The irradiation penetration is typically 1.7cm in muscle and 10cm in fatty tissue, so the heating process can penetrate the breast fairly deeply. In order to reduce excessive heating of the subcutaneous fat layer an active convective cooling of the skin is necessary. The use of this cooling means that the ambient temperature control is more relax in this application than in the previous study. After the stress there is a temperature transient in the breasts. It is the nature of this transient upon which the analysis is based. Thermograms are taken at 30 second intervals for about 8 minutes in order to observe this transient, Figure 28. Tumours appear as hot spots on the recorded thermograms, this is due to differences in the dielectric constants, vascularization, density and specific heat. The result of this process is that small and/or deep tumours can be detected. Furthermore, these infra-red mammograms are easier to interpret than X-ray mammograms; radiographer interpretation of X-ray mammograms is known to highly variable. Another advantage of this approach is that it takes less time to record the series of thermograms, approximately 8 minutes, than the 30 minutes typically required to set up a single conventional thermogram (allowing time for acclimatisation etc.).

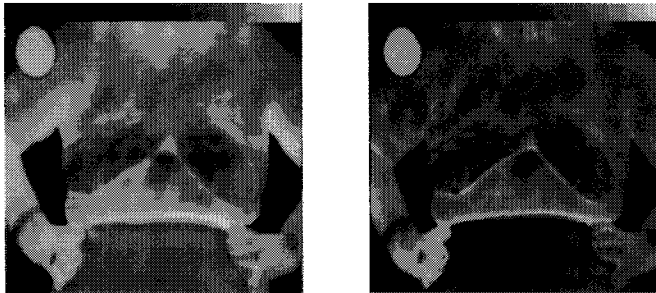


Fig. 28: Infra-red Mammogram Frame 3 and Frame 5

### 7.4 Observed Thermal Characteristics of Breasts

During the course of data collection, some thermal characteristics or behaviour patterns were observed. It is these patterns that the system must be capable of analysing, extracting and quantifying [50].

- **Normal tissues** - Temperature does not rise as high as the abnormal tissues. Temperature drops linearly with time
- **Veins** - Temperature rises slightly higher than the abnormal tissues. Temperature drops more rapidly with time

- **Nipple** - Temperature does not rise as high as the abnormal tissues. Temperature drops quickly with time.
- **Tumours** - Temperature is generally 0.3°C - 1.5°C higher than normal tissues. Temperature remains high but sometimes drops a little.

### **7.5 Experimental Results**

Initial results from this work showed that in some severe cases the ‘hot spots’ (*tumours*) could be identified easily using a colour coded composite eigen images from the Karhunen-Loeve transformation. Also, that the vascular pattern can be identified to a certain extent, this can also help diagnosis because it indicates if micro-calcification exists. However, this approach can only be used as a supplementary diagnostic for breast cancer detection and the following-up treatment. At present X-ray mammography still must be used as the main diagnostic aid. It must be stressed that these are only experimental results. Before such a technique can be applied in practice a far larger evaluation would be required, c.f. the 30 patients used in this study.

## **8. Other Medical Infra-red Applications**

Infra-red thermograms can be used very effectively to assess the depth of a burn. Often the burns are examined by a clinician to decide the degree of burn and the need of surgery [50]. It is also useful in the detection of melanoma as tumorous areas will have a higher temperature than surrounding normal tissues.

Infra-red imaging is widely used in post-operative monitoring. One of the main concerns after an operation is the patient's healing progress and complications such as an infection. After plastic or reconstructive surgery the clinician is concerned with how well the biomaterial is being accepted by the patient's body. Thermal images can help identify areas of inflammation or infection.

## **9. Other Infra-red Applications**

Infra-red technology is used in many application domains other than defence and medicine, for example: infra-red security systems to detect intruders through their bodies' heat and movement; fire fighting; rescue and driving aids.

It is also widely used for machine condition monitoring. The main advantages apart from those mentioned in the above sections are:

- It is immune to electromagnetic noise
- It can be used in explosive environments
- It is conducted in real-time and is only limited by the processing time

### **9.1 Fire Fighting**

In fire fighting airborne imaging systems often provide high-resolution data in a more timely fashion than can space-based systems. Aircraft with on-board infrared and thermal multi-spectral instruments are used for example in fighting

forest fires to acquire images and data on the fire's perimeter, hot spots and direction. Information from this type of imagery can help fire fighters to suppress the fire significantly more efficiently than through the use of space-based imaging systems. This type of infra-red and multi-spectral thermal imaging system can also be applied in law enforcement (e.g. marijuana detection) and forestry (e.g. identification of diseased trees within a forest). Often the collected images or video sequence of images acquired at the site are transmitted to a home base. Image processing, computer vision, pattern processing techniques etc. are then used to analyse the images to develop strategies and action plans.

### ***9.2 Monitoring Radioactive Waste***

One way of dealing with radioactive waste is to pour a mixture of molten glass and waste into canisters for storage. An infra-red detection system has been used to monitor the mixture level inside the canisters during filling, thus prevent spills. The system feeds live thermal video to a remote control room where operators can monitor the hot-glass level during the filling operation. The main advantage of using infra-red technology in this application is that it does not require radiation reference sources. This allows non-radioactive start-up testing of the process and facilitates safe worker entry into the vitrification cell before actually processing the radioactive waste.

## **10. Future Infra-red Sensor Technology**

Most current thermal imaging systems require complex cooling techniques with concomitant penalties in system size and weight and a significant logistic burden. Therefore, much research focuses on un-cooled, compact, low power and cheaper solutions. Ferroelectric detectors are the most promising devices for un-cooled or ambient temperature IR imaging. Good performance can be achieved with large arrays where there is a single ferroelectric detector element for each image pixel. The performance can be further enhanced by reductions in detector noise bandwidth through advanced element and integrated circuit readout design. For the present however, cooled technology is still required for applications which need the highest performance.

A major problem associated with current infra-red sensors is the non-uniformity inherent in the detector array. This requires correction prior to any subsequent image processing. At present correction is achieved using off-focal plane and processing electronics. However, the continuing advances in silicon integrated circuit technology will now allow more functionality, including non-uniformity correction, to be included within each pixel of the focal plane array. Transferring this function onto the focal plane will result in more cost effective solutions, giving improved performance and reliability together with reduced size and weight. In addition, the background pedestal current can be removed, on a pixel by pixel basis, resulting in systems benefits such as improved range and image quality. Moreover, there are opportunities to implement novel and

advanced image processing and pattern recognition techniques on the focal plane array. This will result in new capabilities such as motion detection, clutter rejection, etc. which will significantly enhance the performance of IR systems.

The third generation of infra-red thermal imaging is aiming at large two-dimensional detector arrays, higher operating temperatures, smart focal plane processing and multi-band operation to increase detection and identification performance.

## 11. Acknowledgements

© Copyright QinetiQ Ltd 2003.

## 12. References

- [1] W. L. Wolfe and G. J Zissis (eds), *The Infra-red Handbook*, revised edition, 3<sup>rd</sup> printing. The Infrared Information Analysis (IRIA) Center, Environmental Research Institute of Michigan, USA, 1989.
- [2] A. Rogalski, *Infrared Detectors*, Amsterdam, Gordon and Breach Science Publishers, 2000.
- [3] E. L. Dereniak and G. D. Boreman, *Infrared Detectors and System*, New York, Wiley 1996.
- [4] L. F. Pau and M. Y. Ei Nahas, *An introduction to infra-red image acquisition and classification* (Research Studies Press Ltd. Letchworth, Hertfordshire, England, 1983)
- [5] M. J. Varga & P. G. Ducksbury, *Infra-red Thermography: Techniques and Application*, *Handbook of Pattern Recognition and Computer Vision* (2<sup>nd</sup> Edition) pp891-924. Eds. C. H. Chen, L. F. Pau and P.S. P. Wang, World Scientific Publishing Company, 1998.
- [6] T. Ashley, R. A. Ballingall, J. E. P. Beale, I. D. Blenkinsop, T. M. Burke, J. H. Firkin, D. J. Hall, L. Hipwood, C. Hollier, P. Knowles, D. J. Lees, J. C. Little, and C. Shaw, *Large Format MWIR Focal Plane Arrays*. SPIE Optical and Science Conference, July, 2002.
- [7] R. A. Ballingall, I. D. Blenkinsop, I.M. Baker and Parsons, *Practical Design Consideration in Achieving High Performance from Infrared Hybrid Focal Plane Arrays*, SPIE Vol. 819 *Infrared Technology* (XIII), pp239-249. 1987.
- [8] A. J. Myatt, D. A. Spragg, R. A. Ballingall and I. D. Blenkinsop, *Flexible electronic Control and Correction System for use with IR Focal Plane Arrays.*, SPIE Vol. 891 *Infrared Technology* XIII, 1987, pp 239- 249
- [9] P.W. Foulkes, *Towards Infrared Image Understanding*, Ph.D. Thesis, Engineering Department, Oxford University, 1991.
- [10] J. Pearl, *Fusion, Propagation, and structuring in Belief Networks*, *Artificial Intelligence* 2, pp 241 - 288, 1986.
- [11] P. G. Ducksbury, *Parallel texture region segmentation using a Pearl Bayes Network*, *British Machine Vision Conf.*, Surrey University, pp187-196, 1993.
- [12] P. A. Devijver, *Real-time modelling of image sequences: Based on Hidden Markov Mesh Random Field Models in Decision making in context*, a Course on Statistical Pattern Recognition by P.A. Devijver and J. Kittler, Surrey University, 1989.

- [13] P. G. Ducksbury P, Evidential Reasoning - A Review and Demonstration., UK DTI IED project 1936 : Vision by Associative Reasoning, Report no. VAR-TR-RSRE-92-4, July 1992.
- [14] P. G. Ducksbury, Derivable region detection using a Pearl Bayes Network, IEE Colloquium on Image Processing for Transport Applications, London, 9-Dec, 1993.
- [15] R. W. M. Smith, Conceptual Hierarchical Image Processor (CHIP) : System Design, Issue 1.0, CSE1 Division, DRA Malvern, October 1992.
- [16] D M. Booth D.M. and C. J. Radford, The detection of features of interest in surveillance imagery: Techniques Evaluation, DRA Milestone Report T02C.05.01/26, November 1992.
- [17] N. I. Justisson, Median filtering: statistical properties, in T. S. Huang (ed.), Two dimensional digital signal processing II, Top. Appl. Phys. (Springer Verlag, Berlin, 1981), 161-196.
- [18] A. Rosenfeld and A. C. Kak, Digital Picture Processing, vols. 1 and 2 (Academic Press, New York, 1982).
- [19] Maragos P., R.W., Morphological Systems for Multidimensional Signal Processing, Proc. of IEEE, vol. 78, no 4, April 1990.
- [20] Hussain Z., Digital Image Processing: Practical Applications of Parallel Processing Techniques, Ellis Horwood, 1991.
- [21] Haralick R.M., Mathematical Morphology and Computer Vision, 22<sup>nd</sup> Asilamer Conference. on signals and computers, Pacific Grove, CA, US, 31-Oct/2-Nov 88.
- [22] Sun F.K., Rubin S.L., Algorithm development for autonomous image analysis based on mathematical morphology, Proceedings of SPIE 845: Visual Communications and Image Processing II, 1987.
- [23] J. Serra and P. Solie (eds.), Computational Imaging and Vision, Mathematical Morphology and its Applications to Image ProcessingII, 1987.
- [24] Smith R.W.M., Radford C.J., Development of a connected component labeller DSP module for CHIP, DRA Milestone Report DRA/CIS(SE1)/93.85.04.107a/1.0, October 1993.
- [25] Blackman S.S., Multiple Target Tracking with Radar Applications, Artech House, 1986.
- [26] Brown G., Smith R.W.M., Radford C.J., Target Acquisition and Tracking of a Staring Array Sequence on CHIP, DRA Milestone Report DRA/CIS(SE1)/93.85.04.105, September 1993.
- [27] Kalman, Bucy, New results in Linear Prediction and Filtering, Journal of Basic Engineering, vol. 83-D, pp 95-108, 1961.
- [28] Salmond D.J., The Kalman Filter, The  $\alpha$ - $\beta$  filter and smoothing filters, Royal Aircraft Establishment, report TM-AW-48, February 1981.
- [29] Bridgewater A.W., Analysis of 2<sup>nd</sup> and 3<sup>rd</sup> order steady state tracking filters, AGARD Conference Proceedings no 252, 1978.
- [30] Gulyaev Yu. V, Marov, Koreneva, L..G. and Zakharov PV, Dynamic Infrared Thermography in Humans, IEEE Engineering in Medicine and Biology, November/December, 1995, pp766-771.
- [31] Goin J.E. & Haberman J.D., Automated Breast Cancer detection by Thermography: Performance Goal and Diagnostic Feature Identification. Pattern Recognition Vol. 16, No.2 pp 125 -129, 1983.

- [32] A. Zonka, S. Sideman, O. Verbitsky and R. Beyar, Dynamic Thermography, Analysis of Hand Temperature During Exercise, *Annals of Biomedical Engineering*, Vol. 26, pp 993-998, 1998.
- [33] M. J. Loh ('nee Varga), Application of Statistical Pattern Recognition Techniques to Analysis of Thermograms, 1986, Department of Community Medicine, Cambridge, 1986.
- [34] Lawson R.N, Implications of Surface Temperatures in the Diagnosis of Breast Cancer. *Can. Med. Assoc. J.* Vol. 75, pp 309- 301, 1956.
- [35] Phillips B.H. & Lloyd-Williams K, The Clinical Use of Thermography. *Brit. J. Hosp. Med. Equip. Suppl.*, 1974.
- [36] Carter L.M., The Clinical Role of Thermography. *J Med. Eng. Tech.* Vol. 2 p 125., 1978.
- [37] Kyle M.V., Parr G. Sallisbury R., Page-Thomas P & Hazelman B.L., Prostaglandin E1 Vasospastic Disease and Thermography. Presented at Heberden Round, 1982.
- [38] Rajapakse C., Grennan D.M., Jones C., Wilkinson L. and Jayson M., Thermography in the assessment of peripheral joint inflammation — a re-evaluation. *Rheumatology and Rehabilitation* Vol. 20, pp 81 -87, 1981.
- [39] Varga M.J. & Hanka, R, Dynamic Elastic Image Stretching Applied to Thermographic Images, *Special Issues: IEE Proceedings-I, Communications, Speech & Vision*, 1990.
- [40] B. Widrow, The rubber mask technique, Parts I and II, *Pattern Recognition*, 5, pp175-211, 1975.
- [41] D. J. Burr A Technique for Comparing Curves, *IEEE Proceedings of Computer Society on Pattern Recognition and Image Processing*, pp27-277, 1979
- [42] D. J. Burr' Elastic Matching of Line Drawings, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-3, 6, pp708-715, 1981.
- [43] Kittler J & Devijver P, *Pattern Recognition: A Statistical Approach*, Prentice-Hall International, 1982.
- [44] Kittler J., *Mathematical Methods for Feature Selection in Pattern Recognition.* *Internat. J. Man. Mach. Stud.* Vol. 7. pp 603 - 637, 1975.
- [45] P. E. Hart, The condensed nearest neighbour rule (CNN), *IEEE Transaction on Information Theory*, 14, pp515-516, 1968.
- [46] D. Fogel, E Wasson, E. Boughton, V Proto and P. Angeline, Linear and Neural Models for Classifying Breast Masses, *IEEE Transactions on Medical Imaging*, Vol. 17 no. 3 pp 485-486, 1998.
- [47] NHS Breast Screening Programme, Pathology reporting in breast cancer screening, 2nd ed., Sheffield, The Programme, 1997 Feb. NHSBSP Pub. No. 3.
- [48] J. F. Head, C. A. Lipari and R. L. Elliot Computerized Image analysis of digitised infrared images of the breast from a scanning infrared imaging system. *SPIE Conference on Infrared Technology and Applications XXIC*, Vol. 3436, pp290-294, 2000.
- [49] Ariel I.M. & Cleary J.B., *Breast Cancer Diagnosis and Treatment*, McGraw Hill Book Company, 1987.
- [50] M. J Varga and P. De Muynck, Thermal Analysis and Infra-red Mammography, 11<sup>th</sup> Int. Conference Pattern Recognition, The Hague The Netherlands, 1992.



This page is intentionally left blank

## CHAPTER 26

# HYPERSPECTRAL IMAGING ANALYSIS AND APPLICATIONS

Su May Hsu, Hsiao-hua Burke and Michael K. Griffin

*MIT Lincoln Laboratory  
244 Wood Street, Lexington, MA 02420-9185  
Email: sumayhsu@ll.mit.edu*

Hyperspectral sensing technology has advanced in recent years. There are a number of experimental aircraft platforms that routinely collect hyperspectral-imaging (HSI) data both for civilian and military applications. Space borne data have also become a reality. The realization of utility of HSI data requires coordinated activities by both the development and user communities. In this chapter, examples of hyperspectral application scenarios are discussed. The principle focus is terrain characterization and object detection. Additionally, examples of HSI fusion with other sensors, such as synthetic aperture radar (SAR) and panchromatic imagery, will also be shown to demonstrate the effect of sensor fusion in target detections.

### 1. Introduction

Through the use of dispersive media (prisms, gratings, etc.) or through transform techniques (Fourier, Sagnac, etc.), one can measure a contiguous spectrum with narrow ( $\lambda/\Delta\lambda \sim 100$ ) spectral bands for each pixel in an image and generate a *hypercube*. Covering the VNIR/SWIR spectral region (0.4 to 2.5  $\mu\text{m}$ ), both the NASA Airborne Visible-InfraRed Imaging Spectrometer (AVIRIS)<sup>a</sup> and the DoD HYperspectral Digital Imagery Collection Experiment (HYDICE)<sup>b</sup> sensors have collected a significant amount of data over a variety of military and civilian target scenes<sup>1,2</sup>. A sample scene collected by HYDICE is shown in Figure 1. NASA EO-1, launched in November 2000, is a spaceborne platform providing simultaneous HSI (Hyperion), MSI and panchromatic (ALI)

---

<sup>a</sup> AVIRIS contains 224 different detectors, each with a spectral bandwidth of approximately 10 nanometers (nm), allowing it to cover the entire range between 380 nm and 2500 nm. AVIRIS uses a scanning mirror to sweep back and forth in a whiskbroom fashion, producing 614 pixels for each scan. Each pixel produced by the instrument covers a 20x20 meter square area on the ground (with some overlap of pixels), yielding a ground swath width of approximately 10 kilometers for a flight altitude of 20 km.

<sup>b</sup> HYDICE is push broom, imaging spectroradiometer. HYDICE has a 2-d focal plane array of 320x210. The second dimension of the array records in 210 spectral bands, approximately 10 nm of bandwidth each, covering the range from 400 nm to 2500 nm. Each pixel in the cross-track dimension covers a ground distance of 3 meters, yielding a ground swath width of approximately 1 kilometer for a flight altitude of 6 km.

data<sup>c</sup>. Hyperspectral systems in the LWIR (8-14  $\mu\text{m}$ ) are less mature due to challenges of focal plane and cooling technologies. Nonetheless, it has been demonstrated that LWIR HSI sensors can provide additional capability such as sensing trace amounts of gaseous emissions and distinguishing some unique geologic features. It also offers the day/night operational capability. For the rest of the chapter, however, only HSI in the reflective solar region (0.4 to 2.5  $\mu\text{m}$  or known as VNIR/SWIR, visible-near IR-shortwave IR) will be discussed.

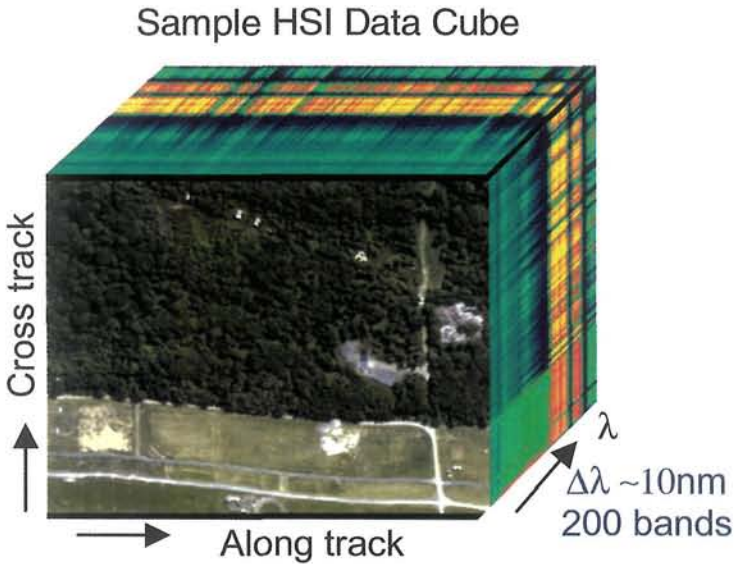


Figure 1. Example HSI data cube collected by HYDICE. Through the use of dispersive media (prisms, gratings, etc.) or through transform techniques (Fourier, Sagnac, etc.), one can measure a contiguous spectrum with narrow ( $\lambda/\Delta\lambda \sim 100$ ) spectral bands for each pixel in an image and generate a *hypercube*.

Currently hyperspectral imaging sensors are routinely used for research and development and in some semi-operational scenarios. HSI offers unique applications such as terrain delimitation, object detection, material identification, and atmospheric characterization. Figure 2 illustrates a taxonomy of HSI applications. Recent processing advances, via hardware and software, as well as the development of automated target detection algorithms, have progressed significantly so that near real-time applications are also within reach.

<sup>c</sup> EO-1 (Earth-Observing 1) is part of NASA's New Millennium Program. The EO-1 spacecraft was launched on November 21, 2000. The instrument payloads on board are Hyperion (HSI), ALI (Advanced Land Imager, MSI and Panchromatic) and AC (atmospheric corrector). Hyperion is a pushbroom, imaging spectrometer with 256 detectors in cross-track, each pixel covers an area of 30x30 meter square on the ground, and a complete spectrum from 400 to 2500 nm in 220 bands. It yields a ground swath width of approximately 7.65 kilometer for a 705 km orbit.

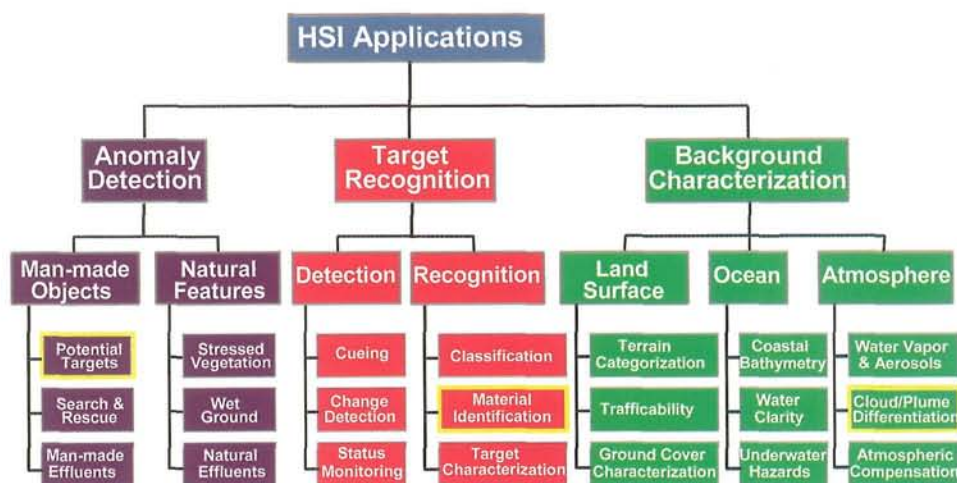


Figure 2. A taxonomy of HSI applications. HSI offers unique applications such as terrain delimitation, object detection, material identification, and atmospheric characterization. The highlighted boxes are discussed in the section of HSI Application Examples.

A passive HSI sensor does not effectively penetrate surface, however, making targets under camouflage and tree canopies difficult to detect. Additionally, to maintain required SNR, HSI spatial resolutions are in general coarser than broadband imagery resolutions. There is great potential to enhance the overall HSI system performance if fusion with other remotely sensed data could be accomplished. For example, in the application of counter camouflage, concealment and deception (CC&D)<sup>d</sup>, HSI can be used to identify ground coverage and surface material and a Foliage Penetration Synthetic Aperture Radar (FOPEN SAR) can determine if any threat objects are under concealment. In the surface surveillance application, HSI can be augmented with High-Resolution broadband panchromatic Imagery (HRI) to optimize simultaneous high spectral and spatial information for enhanced target detection and identification.

The following sections give an overview of HSI processing algorithms, then describe a number of hyperspectral imaging applications, and finally illustrate examples of HSI fusion with other sensors.

## 2. HSI Algorithms

HSI systems produce large data sets that are not easily interpretable by visual analysis and therefore require automated processing algorithms. A wide range of algorithms can be applied for HSI data analysis<sup>3,4</sup>. Figure 3 illustrates a taxonomy of HSI algorithms. In this section, sample algorithms of principal component analysis (PCA), anomaly detection (RX filtering), matched filtering (CEM) and iso-data clustering will be illustrated.

<sup>d</sup> Counter Camouflage, Concealment and Deception (Counter CC&D) is the ability to detect moving and stationary obscured targets in foliage, under camouflage or in shallow hide, and those utilizing deception techniques.



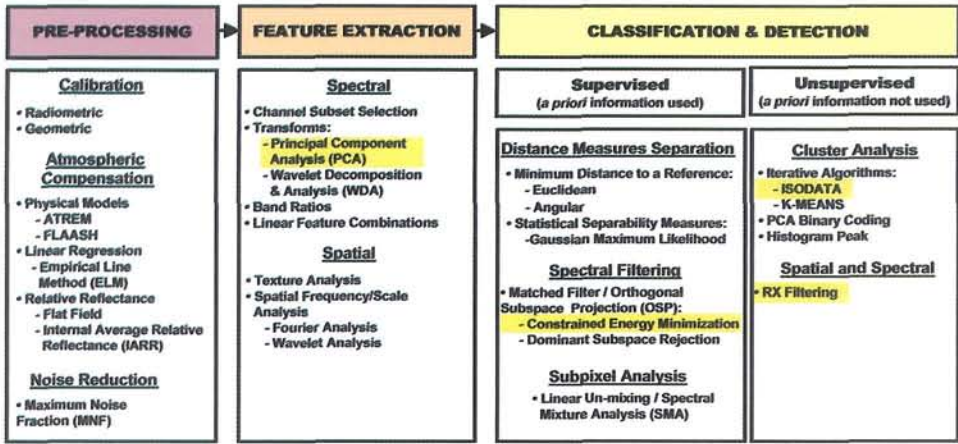


Figure 3. A taxonomy of HSI algorithms of preprocessing, feature extraction and classification. The highlighted ones are discussed in this section.

### 2.1 Principal Component Analysis (PCA)

To reduce the HSI data dimensionality and therefore the computational complexity, feature extraction can be performed on the spectral data before application of image pixel clustering. Principal component analysis is used to de-correlate data and maximize the information content in a reduced number of features<sup>5,6</sup>. The covariance matrix is first computed over the pixel spectra contained in the HSI data cube of interest. Eigenvalues and eigenvectors are then obtained for the covariance matrix  $\Sigma$  as given below:

$$\Sigma = E \{ (X - X_m)(X - X_m)^T \} = \Phi \Lambda \Phi^T,$$

$X$  represents the spectral vector data;  $X_m$  the mean spectral vector over the data cube and  $E$  the average operator over the entire data cube.  $\Phi$  is a matrix consisting of columns of eigenvectors and  $\Lambda$  is a diagonal matrix of eigenvalues.

Using the eigenvectors as a new coordinate system, the HSI data cube is then transformed into principal components called eigenimages. The eigenimages associated with large eigenvalues contain most of the image information while the eigenimages associated with small eigenvalues are noise-dominated. Principal component transform allows determination of the inherent dimensionality and segregation of noise components of the HSI data. The components are ranked in descending order of the eigenvalues (image variances). Since backgrounds constitute the majority of information in the scene, they are contained in the first few principal components. Anomalies, which comprise only a small fraction of the scene, are in higher numbered principal components. Figure 4 displays the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> principal components of a sample HYDICE image. The first component shows the overall intensities of features such as bright treetop and dark tree shade and roads. Area of vegetation is apparent in the 2<sup>nd</sup> component. In the 3<sup>rd</sup> component, image pixels of vehicles and roads are in contrast to the image backgrounds.

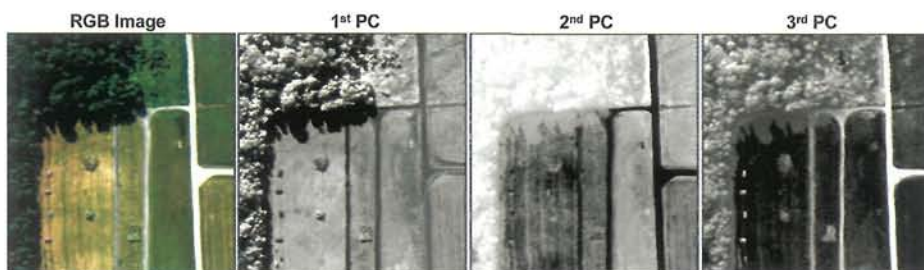


Figure 4. The first three principal components of a sample HSI image. The first component shows the overall intensities of features such as bright treetop and dark tree shade and roads. Area of vegetation is apparent in the 2<sup>nd</sup> component. In the 3<sup>rd</sup> component, image pixels of vehicles and roads are in contrast to the image backgrounds.

## 2.2 Iso-data Clustering

Iso-data clustering, also called iterative migrating means clustering, is an iterative process to cluster image pixels into classes<sup>7,8,9</sup>. A flow diagram of the algorithm is depicted in Figure 5. A number of class centers are first initialized in the feature space. The feature space may be the original spectral space or a transformed space in which spectral feature extraction have been performed such as principal component analysis. An image pixel is

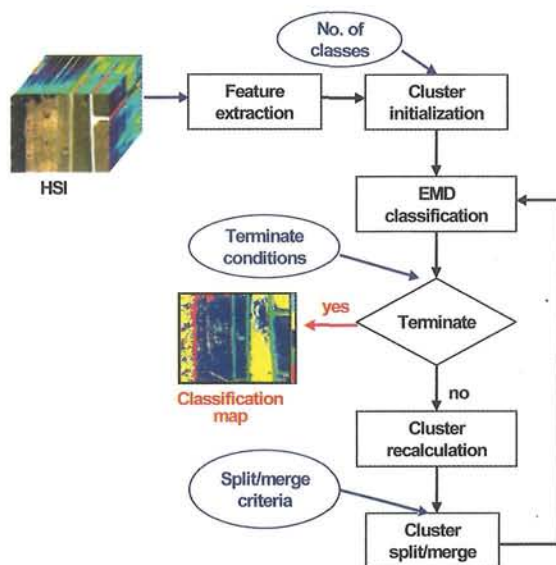


Figure 5. Flow chart of iso-data clustering. A number of class centers are first initialized in the feature space. An image pixel is assigned to the class with minimum Euclidean distance. When all image pixels are classified, the class centers are recalculated. Using the new class centers, the image is reclassified and compared with the classification in the previous iteration. The process is terminated if the changes in classification are less than a threshold or a maximum number of iterations have been reached. Otherwise, the class centers are recalculated for further iterations.

assigned to the class with minimum Euclidean distance. When all image pixels are classified, the class centers are recalculated. The class size and distribution are examined. Closely separated classes are merged and widely distributed classes are split according to specific input criteria. Using the new class centers, the image is reclassified and compared with the classification in the previous iteration. The process is terminated if the changes in classification are less than a threshold or a maximum number of iterations have been reached. Otherwise, the class centers are recalculated for further iterations.

The algorithm gives spectral classifications by iterative convergence without a priori knowledge of the classes in the image. It is conceivable that the process can be extremely slow to converge in a large dimensional space such as the spectral space of HSI data. In practice, the iso-data clustering is often applied on a selected number of spectral features such as a subset of principal components.

### 2.3 RX Filtering

RX filtering is a spatial-spectral processing algorithm for anomaly detection<sup>10</sup>. A spatially moving window is used to calculate local background mean and covariance. The RX filtered value at the center of the window is detected on differences from the local background. The RX filtered value is calculated as the following:

$$RX = (\mathbf{x} - \mathbf{m})^T \Sigma^{-1} (\mathbf{x} - \mathbf{m})$$

**x:** Data spectrum

**m:** Local background mean

**Σ:** Local background covariance

Since large covariance matrix inversion is required at each image pixel, direct implementation of RX filtering on HSI data may be impractical. In practice, the RX filtering is also applied on a selected number of spectral features such as a subset of principal components.

### 2.4 Constrained Energy Minimization (CEM)

CEM, a type of matched filtering, is a spectral processing algorithm for object detection and material identification<sup>11,12</sup>. The CEM value of a spectral vector  $\mathbf{X}$  is calculated as:

$$(\mathbf{X} - \mathbf{m})^T \Sigma_b^{-1} (\mathbf{M} - \mathbf{m}) / (\mathbf{M} - \mathbf{m})^T \Sigma_b^{-1} (\mathbf{M} - \mathbf{m}),$$

where  $\mathbf{M}$  is the reference material spectrum,  $\mathbf{m}$  and  $\Sigma_b$  represent the background mean and covariance matrix. The vector " $\Sigma_b^{-1} (\mathbf{M} - \mathbf{m}) / (\mathbf{M} - \mathbf{m})^T \Sigma_b^{-1} (\mathbf{M} - \mathbf{m})$ " in the formula is the CEM filter. The matched filtering attempts to maximize the energy in the test data associated with the reference spectrum with the constraint of whitening the background clutter. The CEM value is indicative of the fraction content of the reference material in the spectral data  $\mathbf{X}$ .

### 3. HSI Application Examples

To exploit the full value of HSI data in many applications (e.g., identify material, detect unresolved targets), one needs to map the data as measured by the sensor to the same units as the library spectra against which it is compared. This is typically accomplished by performing atmospheric compensation to convert the sensor data to units of surface reflectance<sup>13</sup>. As will be discussed later, this also leads to another application: simultaneous atmospheric characterization. In addition to robust and efficient detection and classification algorithms, a robust and comprehensive material reflectance/emissivity database must be developed.

In this section, key applications for HSI data such as anomaly detection, material identification and background classification will be illustrated with examples followed by a discussion of atmospheric characterization.

#### 3.1 Anomaly Detection

An example of anomaly detection is illustrated in Figure 6 as the result of RX filtering on a HYDICE image. The left panel shows the RGB image of a forest scene. Several vehicles were laid out in the open area. The vehicles are approximately 5-pixel x 11-pixel in size. The RX filtering is implemented using a 21x21 spatial window on four principal components. Due to the rich spectral information, the vehicles are detected by differences from the local background. The right panel of Figure 6 shows detection of the vehicles. Note that other objects in the scene are either not detected or only partly detected due to their larger sizes than the processing window.

#### 3.2 Material Identification

For target detection and material identification, the HSI radiance data are converted to reflectance via atmospheric compensation. The transformation to reflectance allowed the use of matched filtering algorithm in conjunction with material library spectra. Figure 7 shows sample spectral signatures and identification results of fabric net and vehicle paint in the HYDICE forest scene. Material Signatures from a spectral library were used to construct the CEM match filters. The filtered images were thresholded for detection and material identification. In Figure 7, pixels matched with fabric net are marked in red and vehicle paint in green.

#### 3.3 Background Classification

Land-cover characterization and classification is an application of HSI data. Delineation of open area, heavy vegetation of trees and tree lines provides contextual information of the target and background environment. The incorporation of background classification in matched filtering algorithms for material identification has also proven to enhance the filtering performance. Since the background delineation allows for better estimation of background covariance used in the matched filtering algorithms. Figure 8 shows background classification results of the HYDICE forest scene in Figures 6 and 7.



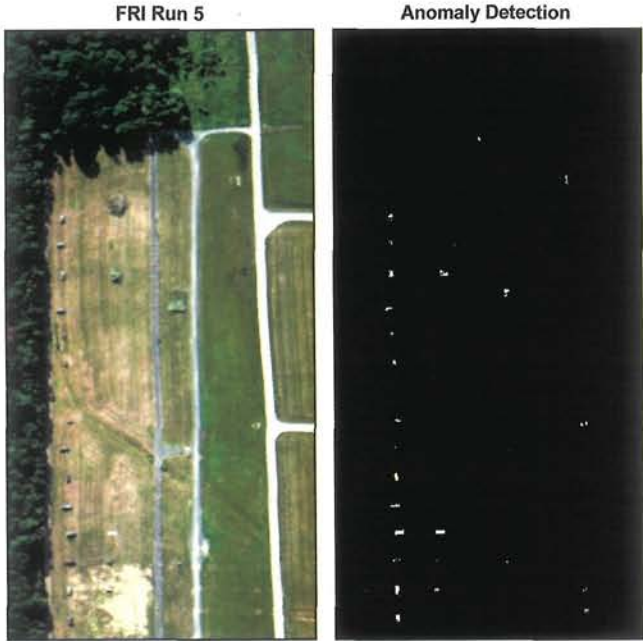


Figure 6. Anomaly detection example. The left panel shows the RGB image of a forest scene. The right panel shows detection of the vehicles with RX filtering. The vehicles are approximately 5-pixel x 11-pixel in size. The RX filtering is implemented using a 21x21 spatial window on four principal components. Due to the rich spectral information, the vehicles are detected by differences from the local background.

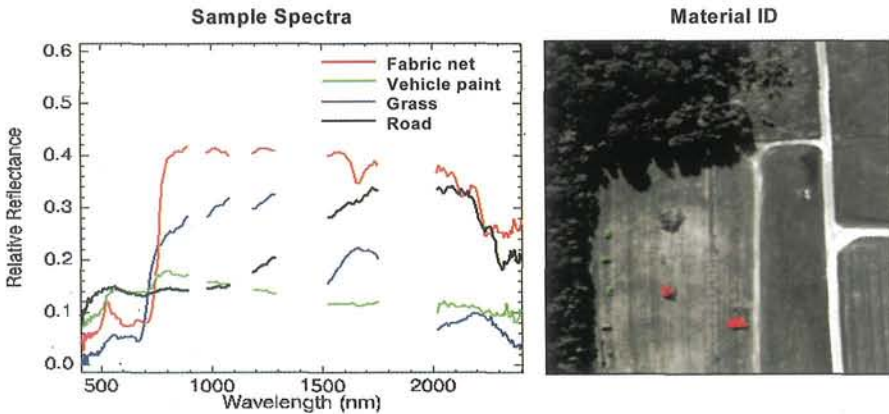


Figure 7. Material identification. Sample spectral signatures are shown on the left and identification results of on the right. Material Signatures from a spectral library were used to construct the CEM match filters and applied to the HYDICE forest scene. The filtered images were thresholded for detection and material identification. In the right panel, pixels matched with fabric net are marked in red and vehicle paint in green.



Figure 8. Example background classification. The left panel is a band image at  $\lambda = 0.55 \mu\text{m}$ . The right image is the background classification of iso-data clustering based on 5 principal components.

Principal component analysis (PCA) was employed for image feature extraction. For the background classification, five principal components were retained. Iso-data clustering divides the image into separate background classes. The eight background classes include road, shade, two ground classes, and several vegetation groups.

### 3.4 Atmospheric Characterization

Conventionally, atmospheric compensation algorithms are first applied to HSI data in order to retrieve surface and target material properties. Unlike conventional EO/IR systems that have to rely on other sources for weather information while data becomes limited during adverse weather, HSI data also allows for atmospheric characterization due to the contiguous spectral coverage. The diagram in Figure 9 illustrates the architecture for theater weather support and delineation of manmade vs. natural obscurants (smoke versus clouds). Certain HSI bands are also transparent to haze and smoke, which allows for battle damage assessment<sup>14</sup>.

The example below illustrates the potential for identifying active fires, smoke plumes and a water cloud in HSI. An AVIRIS scene collected on 20 August 1992 in the foothills east of Linden, CA is employed for illustration of smoke plume and cloud delineation. The scene consists of a grass fire producing a thick plume of smoke extending toward the east (see Fig. 10a). A cloud produced by the thermal properties of the fire overlies the smoke plume. Northwest of the main fire, two smoldering fires are producing a thin veil of smoke that covers much of the upper half of the scene. The southwest portion of the scene is cloud and smoke free; a golf course, lake, roads and rivers can be identified.

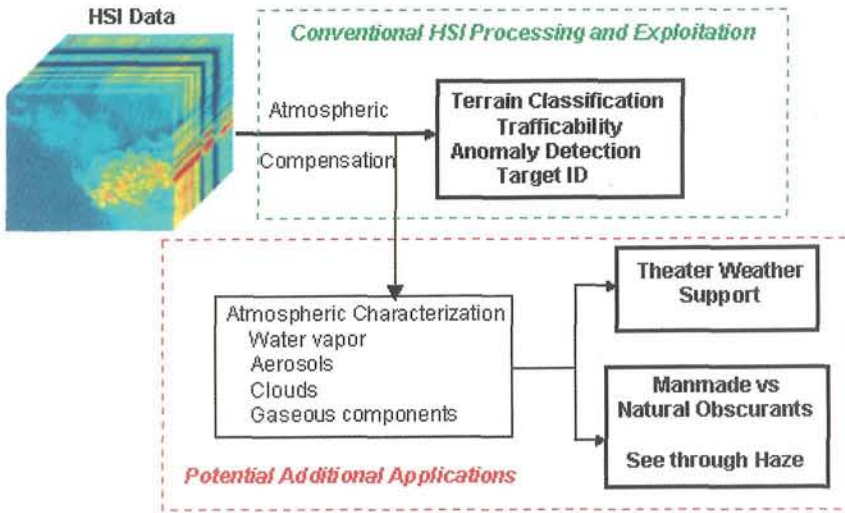


Figure 9. Diagram of architecture for theater weather support and cloud/smoke delineation.

Small shadowed areas are also observed just to the north of the cloud. This scene provides a variety of atmospheric and surface features from which to orient and characterize. A plot of the spectral characteristics of various identified features in the scene is shown in Fig. 10b. The legend displays the user identified features. The cloud is significantly brighter than the smoke over the entire spectral region. The hot area is brightest in the spectral region 2000 to 2500 nm, while the fire pixels are bright for wavelengths greater than 1150 nm. A small dip at 700 nm followed by a rise at 740 nm in the spectrum of large particle smoke indicates its partial transparency to the vegetative background.

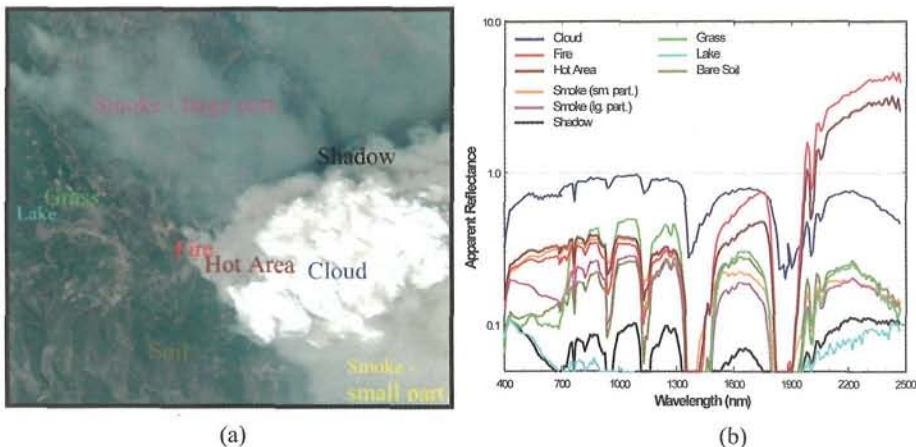


Figure 10. (a) AVIRIS RGB image for the Linden, CA scene collected on 20-Aug-1992, denoting location of various features of interest and (b) a plot of the spectral distribution of the apparent reflectance for those features.



Figure 11 displays the 1<sup>st</sup>, 2<sup>nd</sup> and 5<sup>th</sup> principal components of the AVIRIS data for the Linden Scene. The first component shows the overall intensities of features such as bright clouds and smoke plumes over backgrounds. A dark area that appears to be the source of the thick smoke is apparent in the 2<sup>nd</sup> component. In the 5<sup>th</sup> component, a small fraction of the image pixels are in contrast to the image backgrounds. It is apparent in the PCA that the first two principal components of the Linden AVIRIS data contain background information and the 5<sup>th</sup> component shows an anomaly. A classification can be obtained from these components as shown in Figure 12 with visual identification. All major atmospheric and surface features are identified as to location, extent and type.

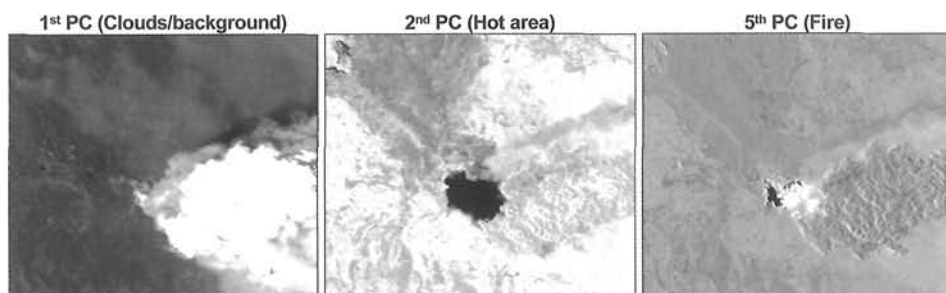


Figure 11. The 1<sup>st</sup>, 2<sup>nd</sup> and 5<sup>th</sup> principal components of AVIRIS data for the Linden scene.

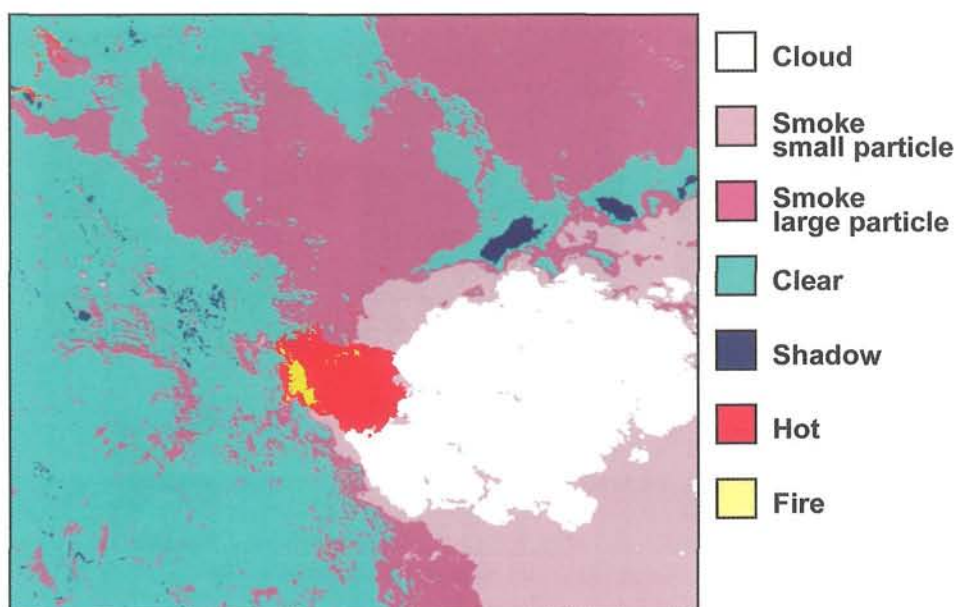


Figure 12. Classification result using the 1<sup>st</sup>, 2<sup>nd</sup> and 5<sup>th</sup> principal components. All major atmospheric and surface features are identified as to location, extent and type.

#### 4. HSI Fusion with Other Sensors

To demonstrate the fusion potential of HSI and other sensors, fusion examples of HSI with SAR and HSI with high-resolution panchromatic imagery (HRI) are shown here<sup>15</sup>.

##### 4.1 HSI and SAR Fusion

There have been many examples in which both SAR and HSI systems collected data in support of military operations. Principles of low frequency SAR and HSI are different and their detection capabilities often complement each other. SAR, an active sensor, penetrates foliage and camouflage nets and detects vehicles but has many false alarms from trees<sup>16</sup>. HSI, on the other hand, does not penetrate surface, but is capable of terrain classification and material identification. A combined SAR and HSI system could have the imaging sensors cue each other in joint observations.

A common data set collected over Vicksburg, Mississippi (Dixie-97, May 1997) is used here to demonstrate the framework of SAR and HSI fusion. Both P-3 UWB (Ultra Wide Band) radar and HYDICE collected data at the target site. Several fabric nets were populated along the tree line around an open area. One fabric net at the tree line covered a vehicle. All other nets were empty or covered non-radar reflecting decoys.

The fusion example is illustrated in Figure 13. SAR and HSI data were first processed separately for detection and terrain classification, respectively. Then co-registration was

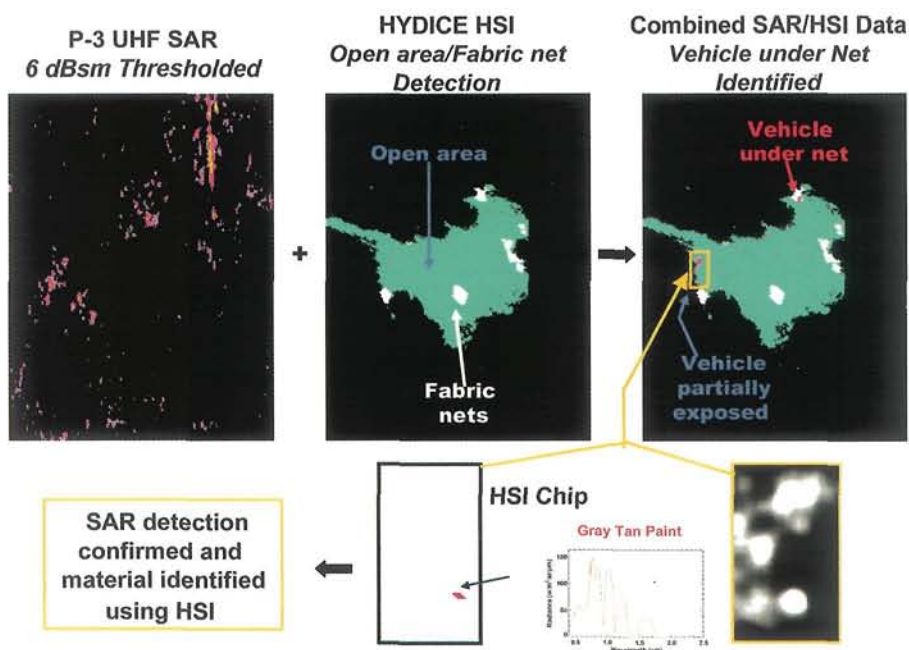


Figure 13. SAR and HSI fusion example. The left panel shows SAR detection; the middle panel shows HSI detection of fabric nets and delineation of open and tree area; the right panel is the combined SAR and HSI results. Terrain mapping reduced SAR false alarm from trees. Detection of concealed targets under nets was verified and detection of partially exposed targets was further confirmed with material identification by HSI.

performed to allow overlay of the images. Terrain mapping reduced SAR false alarm from trees. Detection of concealed targets under nets was verified and detection of partially exposed targets was further confirmed with material identification by HSI. The combined exploitation of HSI with SAR data result in enhanced detection and reduced false alarms.

#### 4.2 HSI with High Resolution Panchromatic Imagery

From a sensor system perspective, integrating HSI with a high-resolution panchromatic imager (HRI) is relatively straightforward and provides synergistic enhancements. While the sensors are similar in spectral coverage, an HRI sensor can offer much finer spatial resolutions than their companion HSI. The rich spectral information from HSI further enhances material identification and target classification. Figure 14 depicts a spatial-spectral analysis approach. Background classification and anomaly detection are first obtained from HSI data. HSI provides background classification and target detection while the HRI provides background and target boundaries with spatial edge detection<sup>17</sup>. These boundaries, combined with results from HSI, spatially enhance the definition of targets and backgrounds. Applied with spectral matched filtering, HSI further identifies the background and target materials.

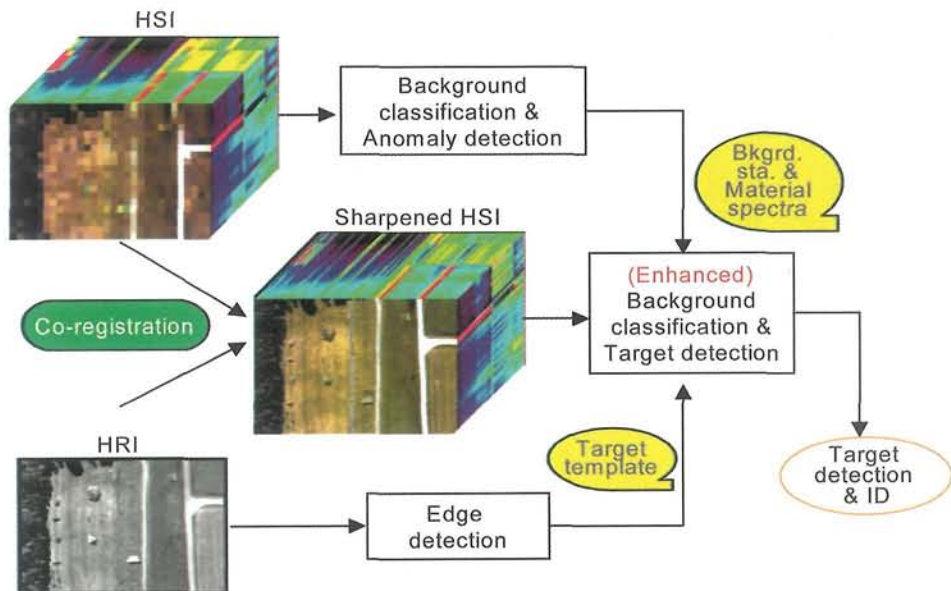


Figure 14. Diagram of a spatial and spectral analysis approach. Background classification and anomaly detection are first obtained from HSI data. HSI provides background classification and target detection while the HRI provides background and target boundaries with spatial edge detection. These boundaries, combined with results from HSI, spatially enhance the definition of targets and backgrounds. Applied with spectral matched filtering, HSI further identifies the background and target materials.



Combined exploitation of HSI and Panchromatic imagery is demonstrated here using the HYDICE Forest data. Data are first generated emulating a high altitude aircraft or a space platform that would provide simultaneous HSI data with a resolution of 4m and panchromatic data with a resolution of 0.8m. The scene consists of several fabric samples and a number of vehicles of two different sizes and two different paint types. Using a combination of image sharpening and spatial/spectral analysis approach results in identification of objects in cued areas<sup>18,19,20,21,22,23,24</sup>. Testing with spectral matched filtering confirms and identifies pieces of fabric and two types of paints on the vehicles in these regions as illustrated in Figure 15. An enlarged view of the vehicle detections is also shown in the figure. The vehicle size and orientation can be determined from the bounding edges. It is classified as large (4x8 m<sup>2</sup>) in size if it is 4 to 7 pixels wide and 8 to 11 pixels long or as small (3x6 m<sup>2</sup>) in size if it is 3 to 5 pixels wide and 6 to 7 pixels long. The colored bar next to each vehicle in the enlarged image depicts its type of paint, size and orientation.

The example of combined HSI and HRI illustrates that in addition to background classification, target size, shape and orientation are also determined; results not possible by either of the sensing modalities alone.

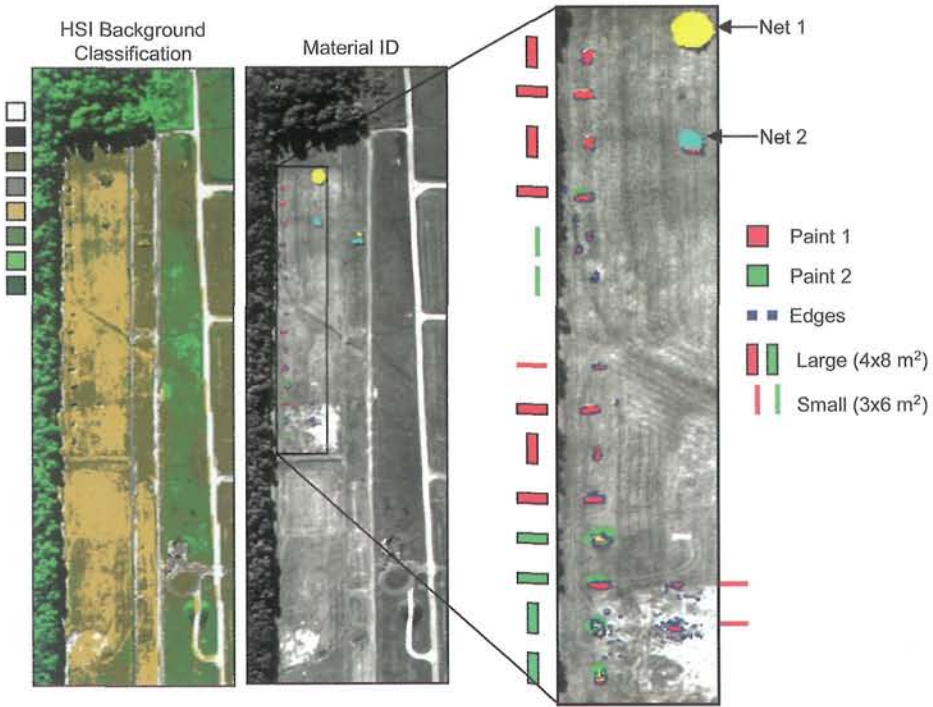


Figure 15. Combined HSI and HRI results of background classification and enhanced spatial and spectral target features. An enlarged view of the vehicle detections is also shown on the right. The vehicle size and orientation can be determined from the bounding edges. The colored bar next to each vehicle in the enlarged image depicts its type of paint, size and orientation.

## 5. Conclusion

In this chapter, the status of VNIR/SWIR hyperspectral sensing and selected algorithms together with several of its applications are discussed. It is important to recognize that the utility of HSI varies depending on the sensor spatial resolutions. For example, a high altitude surveillance sensor with coarse spatial resolution provides functions such as terrain characterization and trafficability analysis. For focused area analysis with high spatial resolution, HSI provides target detection and identification applications.

Perhaps the greatest power of the HSI technology is its combined use with data from other types of sensors. Sensor fusion results demonstrated better context information and false alarm mitigation when used in conjunction with SAR. When combined with higher spatial resolution panchromatic imagery, additional information regarding size, shape, and orientation about surface objects can be derived than those with data from either sensor alone.

## Acknowledgements

The authors are grateful for SITAC for the HYDICE data. This work was conducted under ODUSD S&T Hyperspectral Technology Assessment Program (HTAP) at Lincoln Laboratory.

This work was sponsored by the Department of Defense under contract F19628-00-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and not necessarily endorsed by the United State Government.

## References

1. Vane, G., R.O. Green, T.G. Chrien, H.T. Emmark, E.G., Hansen, and W.M. Porter, "The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)", *Remote Sensing and Environment* 44, pp. 127-143, 1993
2. Rickard, L.J. et al., "HYDICE: An Airborne System for Hyperspectral Imaging", *SPIE Proceedings, Imaging Spectrometry of the Terrestrial Environment*, Vol. 1937, p. 173, 1993
3. Shaw, G. and D. Manolakis, editors, Special Issue on Exploiting Hyperspectral Imagery, *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 12-16, January 2002
4. Manolakis, D., G. Shaw, "Detection Algorithms for Hyperspectral Imaging Applications," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 29-43, January 2002
5. Richards, John, *Remote Sensing Digital Image Analysis*, Springer-Verlag, 1994
6. Geladi, Paul, and Hans Grahn, *Multivariate Image Analysis*, John Wiley & Sons, 1997
7. Sabins, Floyd F. Jr., *Remote Sensing, Principles and Interpretation*, W.H. Freeman and Company, New York, 1978
8. Schott, John R., *Remote Sensing, The Image Chain Approach*, Oxford University Press, 1997
9. Schowengerdt, Robert A., *Remote Sensing, Models and Methods for Image Processing*, Academic Press, 1997



10. Yu, X. and I.S. Reed, "*Comparative performance analysis of adaptive multiband detectors*", IEEE Transactions on Signal Processing, 41:2639-2656, August 1993
11. Harsanyi, Joe. C. and C.I. Chang, "*Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach*", IEEE Transactions on Geoscience and Remote Sensing, V. 32, pp. 779-785, 1994
12. Manolakis, Dimitris, Gary Shaw and Nirmal Keshava, "*Comparative Analysis of Hyperspectral adaptive Matched Filter Detectors*", SPIE Aero Sense, Orlando, Florida, 24-28 April 2000
13. Gao, B.C., K. B. Heiderbrecht and A. F. H. Goetz, "*Atmosphere Removal Program (ATREM) Version 2.0 Users Guide*", Center for the Study of Earth from Space/CIRES, University of Colorado, Boulder, Colorado, 1996
14. Burke, H.K., M.K. Griffin, J.W. Snow, C.A. Upham and C.M. Richard, "*Examples of Atmospheric Characterization Using Hyperspectral Data in the VNIR, SWIR and MWIR*", Proceedings, SPIE AeroSense 2001 Conference
15. Hsu, S.M. and H.K. Burke, "*Potential of Fusion between HSI and other Sensors*", Proceedings, ISSSR, Las Vegas, Nevada, 1-4 November 1999
16. Novak, L.M., G.J. Owirka and C.M. Netishen, "*Performance of a High-Resolution Polarimetric SAR Automatic Target Recognition System*", Linc. Lab. J., Vol 6, 11, 1993
17. Gonzalez, R.C. and P. Wintz, *Digital Image Processing*, Addison-Wesley, 1987
18. Patterson, Tim J. and Michael E. Bullock, "*Radiometrically Correct Sharpening of Multispectral Images using a Panchromatic Image*", p. 252-264, SPIE Vol. 2234
19. Patterson, Tim J., Lester Gong, Robert Haxton and Troy Chinen, "*Extension of Sharpening Techniques to Hyperspectral Data*", p. 114-122, SPIE Vol. 3372, April 1998
20. Vrabel, Jim, "*Advanced Band Sharpening Study*", p. 73-84, SPIE Vol. 3071, April, 1997
21. Goforth, Mark A., "*Multispectral Image Sharpening with Multiresolution Analysis and the MTF*", p. 123-131, SPIE Vol. 3372, April 1998
22. Gross, Harry N. and John R. Schott, "*Application of Spatial Resolution Enhancement and Spectral Mixture Analysis to Hyperspectral Images*", p. 30-41, SPIE Vol. 2821, 1996
23. Gross, Harry N. and John R. Schott, "*Application of Spectral Mixture Analysis and Image Fusion Techniques for Image Sharpening*", p. 85-94, Remote Sensing and Environment 63, 1998
24. Zhang, Bing, Jianguo Liu, Xiangjun Wang and Changshan Wu, "*Study on the Classification of Hyperspectral Data in Urban Area*", p. 169-172, SPIE Vol. 3502

## Subject Index

<b>A</b>		cloud classification	228
adaptive spatial filtering	164	clustering, k-means	42
adaptive Savitzky-Golay filter	487	clustering, model-based	42
Algebraic Reconstruction Technique	264	coastal monitoring	141
anomaly detection	597	coherence	126
aperture response function	261	compression,	
Arctic-ice motion	155	lossless	503
asymmetric Gaussian function	487	near-lossless	503
atmospheric characterization	596	virtually lossless	504
attributed relational graph	38	compression ratio	75, 503
automated scene classification	228	condensed nearest neighbor classifier	585
automatic content extraction	35	connected component labelling	65,
automatic detection of urban regions	558		352, 567
automatic target detection	350	constrained energy minimization	600
automatic target segmentation	349	content-based retrieval	35
		contrast stretch	538
<b>B</b>		<b>D</b>	
background classification	601	database development	242
Bayesian classifier	289	data chain model	4
Bayesian estimation	171	data compression	329, 503
Bayesian fusion	355	data fusion	18, 350
Bayesian wavelet denoising	173	data mining	228
block diagonalisation	14	Dempster-Shafer fusion	349
		depth imaging	469
<b>C</b>		differential PCM	509
change vector analysis	407	digital elevation model (DEM)	35, 96
Chlorophyll-a concentration	158	digital seismic processing	430
classification,		discrete cosine transform	25, 504
cascade	290	discriminant analysis	23
contextual	385, 408	distortion measures	506
maximum likelihood method	365	dual sensors	429
object based	385	dynamic corrective matching	574
partially supervised	285	dynamic deconvolution	429
pixel	315, 385		
shape-based	392	<b>E</b>	
supervised	66, 405	edge-preserving smoothing (EPS)	455
unsupervised	105, 329, 413	Einstein deconvolution	429
classifier architecture	294	empirical orthogonal functions	23
classifier ensemble	318	entropy coding algorithms	509
cloud ceiling height estimation	239		

- entropy regularization 542
- expectation-maximization (EM)  
 algorithm 44, 167, 289, 405
- extrapolation algorithms 259
- F**
- features, 39, 235  
 oceanographic 156  
 pixel level 35  
 region level 35  
 scene level 35  
 spatial 36  
 texture 36, 376
- feature extraction 25, 37, 147, 365
- feature preservation 163
- feature selection algorithms 230, 369
- forward looking IR imagery 564
- fuzzy C-mean algorithm, extended 329
- fuzzy membership function 35
- G**
- Gamma-MAP filter 166
- Gaussian mixture model 43, 172
- Gaussian pyramid 287, 516
- geographical information system 285, 408
- Geostationary Operational Environmental Satellite (GOES) imagery 227
- Grochenig theory 263
- ground penetrating radar (GPR) 535
- GPR images 538
- ground truth 242, 409
- H**
- hierarchical scene modeling 57
- hierarchical segmentation algorithm 64
- high level semantics 39
- higher order statistics 28
- histogram screening for feature detection 159
- Hughes phenomenon 12
- hyperspectral image (HSI) data 329, 521, 595
- HIS data cube 596
- I**
- ice edge tracking 141
- ice floe tracking 141
- image enhancement 27, 141, 260, 535
- image mining 35
- independent component analysis (ICA) 23
- information mining 63, 349
- infrared (IR) imagery 558
- inverse filtering 215
- inverse scattering subseries 469
- irregular sampling reconstruction 255
- iso-data clustering 597
- iterated conditional mode (ICM) 175
- J**
- Jeffret-Matusita distance 365
- joint cumulant ICA algorithm 28
- K**
- K-mean clustering 42, 107, 329
- Kittler-Illingworth thresholding method 410
- Kittler & Young method 585
- k-nearest centroid neighbor 315
- k-nearest neighbor classification 228
- k-surrounding neighbor 317
- knowledge chain model 18
- Knowledge Discovery for Databases (KDD) 227
- L**
- labelling algorithms 7, 63
- land cover map 289, 365
- land cover transition 406
- landmine detection 535
- Landsat image 216
- Landsat Multispectral Scanner (MSS) 7, 63
- Landsat Thematic Mapper (TM) 7, 63, 418
- layered-earth model 442

leave-one-out method	316
Lee filter	164
line recognition	395
L-infinity distance	504

## M

majority voting	303, 312, 356
Markov random field,	36, 171, 329, 407
compound	333
Gaussian	30, 329
material identification	596
maximum <i>a posteriori</i> (MAP)	
estimate	165, 329
maximum homogeneous region	
(MHR) filter	165
median filter	540, 565
meteorological parameters	240
minimum mean square error (MMSE)	
estimate	186
mixed pixels	16
morphological operations	566
multi-look SAR images	166
multiple cascade classifier	293
multiple classifier system	286
multiple sensors	153
multiresolution image analysis	383
multispectral image merging	215
multitemporal image data	285, 406

## N

Nagao's method	455
neural networks,	
back-propagation trained (BP)	366
learning vector quantization	
(LVQ)	365
multi-layered	366, 385
probabilistic (PNN)	408
radial basis function (RBN)	295
self-organizing map (SOM)	365
support vector machine (SVM)	67
nonlinear least square fit	487
normalized difference vegetation	
index (NDVI)	487

## O

ocean remote sensing	102, 141
ocean surface slope estimation	101
oil slicks	147
outline derivation	573

## P

parametric contour tracing	70
Pearl's Bayes network	589
phenological parameters	493
piecewise linear polygonal	
approximation	63
pixel level fusion	349
polarization angle estimation	89
polarization decomposition theorem	109
power line surveillance	383
prediction error filtering	455
predictive deconvolution	431
principal component analysis	23,
	329, 407
probability of detection	359

## R

reference set condensation algorithm	323
reference set reduction algorithm	322
reflection seismic method	429
region relationships	45
remote detection	429
remote sensing,	v, 3
active	6
active microwave	255
passive	6
resolution enhancement	255
RX filtering	600

## S

sampling density	262
scale dependency	169
scattering mechanisms,	
canonical	122
dominant	106
scattering theory	473
scatterometer image reconstruction	
(SIR)	255

scene modeling	35	interferometric data	105
seasonality parameter extraction	487	polarimetric data	89, 366
segmentation,		SAR speckle filter	167
Bayesian	41	<b>T</b>	
range	352	temporal correlation	285
reflectivity	352	terrain delimitation	596
region	561	thermogram	555
shape-based	349	three-dimensional seismic images	455
unsupervised	329	transform methods,	23
seismic data processing	472	Fisher	24, 415
seismic edge-detection algorithm	455	orthogonal	25
seismic ghost reflection elimination	488	orthonormal	503
seismic reflection data	469	scale-space	387
sensor,	3	tropical cyclone intensity estimation	233
altimeters	255	two-dimensional histogram	417
AVIRIS	14, 521, 595	<b>U</b>	
hyperspectral	8	unit-impulse reflection response	429
infrared	555	unsupervised change detection	405
laser radar	349	<b>V</b>	
microwave	255	visual grammar	35
optical	7	<b>W</b>	
radar	10	wavelet coefficients,	170
radiometer	255	thresholding	169
scatterometer	255	wavelet decomposition	167
sensor characteristics	3	wavelet denoising	167
sensor fusion	349, 609	wavelet despeckling	166
sensor technology	4, 590	wavelet reconstruction	209
shape analysis and retrieval	63	wavelet representation,	197
shape similarity measure	77	orthogonal multispectral	197
source decorrelation	509	wavelet transform,	31
spatial resolution	6	discrete	168, 197, 504
speckle noise	163	dyadic	201
speckle reduction	166	Gaussian	141
spectral resolution	4	inverse discrete	169
SPOT Panchromatic image	199	Maxican-hat	157
subsurface geological structure	430	two-dimensional	141
subsurface sensing	535	Wiener filtering	262
support vector machine (SVM)	13, 64	Wishart statistics	107
SVMLB library	68		
Synthetic Aperture Radar (SAR),	89,		
141, 505, 595			
image despeckling	163		