Advanced Sciences and Technologies for Security Applications

Panagiotis Karampelas
Thirimachos Bourlai *Editors*

Surveillance in Action

Technologies for Civilian, Military and Cyber Surveillance

Springer

Advanced Sciences and Technologies for Security Applications

The series Advanced Sciences and Technologies for Security Applications comprises interdisciplinary research covering the theory, foundations and domain-specific topics pertaining to security. Publications within the series are peer-reviewed monographs and edited works in the areas of:

- biological and chemical threat recognition and detection (e.g., biosensors, aerosols, forensics)
- crisis and disaster management
- terrorism
- cyber security and secure information systems (e.g., encryption, optical and photonic systems)
- traditional and non-traditional security
- energy, food and resource security
- economic security and securitization (including associated infrastructures)
- transnational crime
- human security and health security
- social, political and psychological aspects of security
- recognition and identification (e.g., optical imaging, biometrics, authentication and verification)
- smart surveillance systems
- applications of theoretical frameworks and methodologies (e.g., grounded theory, complexity, network sciences, modelling and simulation)

Together, the high-quality contributions to this series provide a cross-disciplinary overview of forefront research endeavours aiming to make the world a safer place.

The editors encourage prospective authors to correspond with them in advance of submitting a manuscript. Submission of manuscripts should be made to the Editor-in-Chief or one of the Editors.

More information about this series at http://www.springer.com/series/5540

Panagiotis Karampelas · Thirimachos Bourlai
Editors

Surveillance in Action

Technologies for Civilian, Military and Cyber Surveillance

Springer

Editors

Panagiotis Karampelas
Department of Informatics and Computers
Hellenic Air Force Academy
Attica
Greece

Thirimachos Bourlai
Multispectral Imagery Lab, Lane
Department of Computer Science and Electrical Engineering
West Virginia University
Morgantown, WV
USA

ISSN 1613-5113 ISSN 2363-9466 (electronic)
Advanced Sciences and Technologies for Security Applications
ISBN 978-3-319-68532-8 ISBN 978-3-319-68533-5 (eBook)
https://doi.org/10.1007/978-3-319-68533-5

Library of Congress Control Number: 2017955255

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The world is facing new challenges in all aspects of business involving the geopolitical and military environment. Religious radicalization, arm races, refugees' movements, global cyberattacks and terrorist attacks are some of these challenges. To be able to operate in such a volatile environment, various government agencies and organizations, including the department of defense, homeland security and intelligent services, as well as surveillance and security businesses, attempt to gain a tactical advantage to mitigate challenges or to make smart decisions, by collecting and processing all sources of important information relevant to their mission. On the one hand, smart phones, smart watches and in general smart devices, including surveillance sensors and on the other online social network platforms, have become the main vehicles to collect surveillance in action information.

Governments, especially after 9/11, started running intensive surveillance programs intending to identify potential terrorist threats. At the same time, companies use data collected through various devices or sensors in order to understand customer behavior, and tune it to improve security and protect their interests. Traditional sources of information are also used by various parties to acquire strategic knowledge against their competitors. As a result, there is an increased need for novel methods of surveillance that can be adapted to the new and dynamic military and civilian environment of the modern world.

In this context, the book attempts to address the aforementioned needs by presenting novel research by different experts around the world in the areas of military, civil, and cyber surveillance. The book is organized into three parts (themes) that present the current trends and novel techniques in the areas of (i) surveillance of human features, (ii) surveillance for security and defense and (iii) cyber surveillance.

In the first part of our book, *Surveillance of Human Features*, the contributors review surveillance systems that use biometric technologies. They propose various novel approaches that cover different topics such as gait recognition, facial soft biometrics, face-based physiology, face recognition using frontal and profile images, cross-spectral iris recognition or examine the facial characteristics in the visible or in different bands and wavelengths of the infrared (IR) spectrum for the purpose of improving recognition performance.

The second part of our book, *Surveillance for Security and Defense*, summarizes mainly surveillance techniques used by the army and secret services. It also discusses the ethical issues raised by the use of surveillance systems in the name of counterterrorism and security. More specifically, the different generations of satellite surveillance systems are presented and the requirements for real-time satellite surveillance for military use are described. The new standards of surveillance using Unmanned Air Vehicles (UAVs) and drones are explored. Then, novel surveillance techniques are proposed in order to detect stealth aircrafts and drones. Due to the increase of cross-border terrorist threats, the book contributors highlight novel techniques for maritime border surveillance, bio-warfare and bioterrorism detection. Next, the way that intelligence services operate and use surveillance in the new era of social media is explored and, finally, the right and conditions under which the governments need to use surveillance technologies is discussed.

The last part of the book is *Cyber Surveillance*. It focuses on a series of computational techniques that can be used for cyber surveillance. First, a review of data hiding techniques that are used to hinder electronic surveillance is provided. Then, novel methods to collect and analyze information by social media sites (such as Twitter or other organizational communication systems) are presented. The focus is to discuss approaches capable to detect inside and outside threats by different individuals such as spammers, cybercriminals, suspicious users or extremists in general. Finally, the third part of our book concludes by examining how high performance environments can be exploited by malicious users and what surveillance methods need to be put in place to protect this valuable infrastructure.

We hope this book can become a reference work for military and law enforcement personnel using surveillance-related technologies, as well as researchers (academic or not), Masters and Ph.D. students who want to focus in the area of surveillance technologies and want to be updated with the current developments in the area of military, civilian, and cyber surveillance. Finally, we would like to thank all the contributors of the book for the high-quality work they have submitted to us and their support in the coordination of this publication.

Dekelia Air Base, Attica, Greece Panagiotis Karampelas
Morgantown, WV, USA Thirimachos Bourlai

Acknowledgements

The process of reviewing and assessing the submitted chapters was very challenging, especially in an edited book in which a wide range of topics is covered. As a result, we kindly requested the assistance of subject matter experts in order to provide their technical expertise to the editors of the book as well as to individual chapter authors. The process itself was intense and thorough and, thus, we would like to express our gratitude to the academics and researchers who devoted their time to provide constructive comments to the authors of the book chapters.

In this respect, we would like to thank:

- Jeremy Dawson, Gianfranco Doretto, Yanfang Ye, Michael Martin, Erin Moore, Saeid Motiian, Neeru Narang from West Virginia University, U.S.A.
- Ioanna Lekea, Panagiotis Papakanellos, Konstantinos Zikidis from the Hellenic Air Force Academy, Greece
- Shuowen (Sean) Hu, Damon Conover, Patrick Rauss and Matthew Thielke from the U.S. Army Research Laboratory, U.S.A.
- Antwan Clark from Johns Hopkins University, U.S.A.
- Rafael Tena from Disney Research, U.S.A.
- Nikolaos Kannelopoulos from the Hellenic Army Academy, Greece
- Yiannis Manoussopoulos from ELGO DEMETER, Plant Protection Division of Patras, Greece
- Manousos Kambouris from the Department of Food Technology, ATEI of Thessaly, Karditsa, Greece
- Vasileios Anastopoulos from the Systems Security Laboratory, Department of Digital Systems, University of Piraeus, Greece
- Kyros Yakinthos from the Aristotle University of Thessaloniki, Greece, and
- Haris Kontoes from the National Observatory of Athens, Greece.

We would also like to thank Annelies Kersbergen, Senior Assistant Editor, NATO/IUC at Springer Physics for her continuous support, valuable feedback and suggestions at all stages of the production of the book.

Contents

Part III Cyber Surveillance

Editors and Contributors

About the Editors

Panagiotis Karampelas holds a Ph.D. in Electronic Engineering from the University of Kent at Canterbury, UK and a Master of Science degree from the Department of Informatics, Kapodistrian University of Athens with specialization in "High Performance Algorithms". He also holds a Bachelor degree in Mathematics from the same University majoring in Applied Mathematics. He has worked for 3 years as an associate researcher in the Foundation for Research and Technology-Hellas (FORTH), Institute of Computer Science. During this collaboration, he had the opportunity to work in numerous European funded research projects in the area of Information Society Technologies Programme practicing his expertise in User Interface Design, Usability Evaluation and Programming. He has also worked for several years as a user interface designer and usability expert in several IT companies designing and implementing large-scale research and commercial information systems. Among others he collaborated with Hellenic Army, General Staff, Signal Directorate as Studies and Research Officer, IntelliSolutions, Greece as Senior Developer implementing large-scale information systems in Vodafone Greece, General Bank of Greece, Microsoft, etc. Then he joined Hellenic American University as Assistant Professor teaching human computer interaction, programming, managing information systems and database management in the Graduate and Undergraduate programs. In parallel, he was visiting the School of Pedagogical and Technological Education, Greece as Assistant Professor teaching courses at the Department of Electrical and Electronic Engineering Educators. Currently, he is with the Department of Informatics and Computers, at the Hellenic Air Force Academy teaching courses to pilots and engineers.

His areas of interest include Human Computer Interaction, Information Visualization, Data Mining, Social Network Analysis, Counterterrorism Informatics, Power Management System, Artificial Neural Networks, Power Transmission and Distribution Systems. He has published a number of books and research articles in his major areas of interests in international journals and conferences. He is the author of the book "Techniques and Tools for Designing an Online Social Network Platform" published in Lecture Notes in Social Networks (2013) and one of the Editors in the book "Electricity Distribution—Intelligent Solutions for Electricity Transmission and Distribution Networks" in the Book Series Energy Systems (2016). He is also a contributor to the Encyclopedia of Social Network Analysis and Mining. He serves as a Series Editor in the Book Series Advanced Sciences and Technologies for Security Applications and as an Associate Editor in the Social Network Analysis and Mining journal. He also serves as program committee member in a large number of scientific journals and international conferences in his fields of interests. He is also member of the Advisory Board and organizer of the European Intelligence and Security Informatics Conference (EISIC) and participates in the Organizing Committee of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

Thirimachos Bourlai is an associate professor in the Lane Department of Computer Science and Engineering at WVU. He also serves as an adjunct assistant professor in the WVU School of Medicine, Department of Ophthalmology, and the Department of Forensic and Investigative Science. He is the founder and director of the Multi-Spectral Imagery Lab at WVU. After earning his Ph.D. in face recognition and completing a post-doctoral appointment at the University of Surrey (U.K.), Bourlai completed a second post-doc in a joint project between Methodist Hospital and the University of Houston, in the fields of thermal imaging and human-based computational physiology. He joined the staff at WVU in 2009 serving as a visiting professor and later as a research assistant professor in the Lane Department.

Bourlai served and has been invited to serve as chair at a number of biometrics conferences including ICB, BTAS, IJCB, FG, SPIE, ISS World Americas, IDGA, ACI and the Biometrics Institute. He has served as a member on technical program committees for other primary computer vision- and biometrics-focused conferences. Several governmental agencies, organizations and academic institutions have invited Bourlai to present his work, including the CIA, NSA, U.S. Secret Service, U.S. Army (various divisions), FBI, Amazon, SRC, Biometrics Institute, NLETS, IDGA, the Biometrics Summit Conference, the IEEE Signal Processing Society, University of Notre Dame, University of Pittsburgh, Rutgers University and the University of Newcastle (UK). He is also a reviewer for a number of premier journals in computer vision, biometrics and related areas (i.e., TPAMI, TIFS, IJCV, TCSVT, PRL, TIP, MVA).

The primary focus of Bourlai's research is on designing and developing technology for supporting, confirming and determining human identity in challenging conditions using primarily face images, captured across the imaging spectrum (including ultraviolet, visible, near-infrared, short-wave IR, mid-wave IR and long-wave IR) and secondarily, other hard or soft biometrics including iris, fingerprints, ears and tattoos. Additionally, he has worked on liveness detection problems using face and pupil dynamics; mobile biometrics; multi-spectral eye and pupil detection; and matching mugshots from government ID documents (e.g. passports or driver's licenses) to live face images, which includes the development of image restoration techniques to recover the original mugshot behind the watermarks. Bourlai has collaborated with experts from academia, industry and the government on projects that involve the collection of various biometric data (including multi-spectral face images, irises and tattoos) and the design and development of biometric- and forensic-based algorithms that can support law enforcement and military operations.

Contributors

Ayman Abaza Cairo University, Cairo, Egypt

Mohammed A. M. Abdullah School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, UK

J.M. Absher University of Maryland College Park, Baltimore, MD, USA

Raid R. Al-Nima School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, UK

Nawaf Yousef Almudhahka University of Southampton, Southampton, UK

Marco Botta Computer Science Department, University of Torino, Turin, Italy

Imed Bouchrika Faculty of Science and Technology, University of Souk Ahras, Souk Ahras, Algeria

Thirimachos Bourlai Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, USA

Davide Cavagnino Computer Science Department, University of Torino, Turin, Italy

Jonathon A. Chambers School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, UK

A.D. Clark Johns Hopkins University, Baltimore, MD, USA

Satnam S. Dlay School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, UK

Susan El-Naggar West Virginia University, Morgantown, USA

Vassiliki Gkantouna Department of Computer Engineering & Informatics, School of Engineering, University of Patras, Rio Campus, Patras, Greece

Ben Harbisher De Montfort University, Leicester, UK

Jonathon S. Hare University of Southampton, Southampton, UK

Lawrence Hornak University of Georgia, Athens, GA, USA

Zafeiria-Marina Ioannou Department of Computer Engineering & Informatics, School of Engineering, University of Patras, Rio Campus, Patras, Greece

Manousos E. Kambouris Department of Pharmacy, University of Patras, Patras, Greece; Department of Food Technology, ATEI of Thessaly, Karditsa, Greece

Charles Kamhoua Cyber Assurance Branch, Air Force Research Lab, Rome, NY, USA

Panagiotis Karampelas Hellenic Air Force Academy, Attica, Greece

Anastasios Kokkalis Hellenic Air Force Academy, Dekelia Air Base, Dekelia, Attica, Greece

Alexandros Kolovos Department of Automatic Control, Aerospace Technology, Defense Systems and Operations, Hellenic Air Force Academy, Dekelia Air Base, Dekelia, Greece

Kevin Kwiat Cyber Assurance Branch, Air Force Research Lab, Rome, NY, USA

Theodore I. Lekas Hellenic Air Force Academy, Dekelia Air Base, Dekelia, Attica, Greece

Ioanna K. Lekea Division of Leadership-Command, Humanities and Physiology, Dekeleia Air Base, Department of Aeronautical Sciences, Hellenic Air Force Academy, Attiki, Athens, Greece

Neeru Narang Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, USA

Mark S. Nixon University of Southampton, Southampton, UK

Laurent Njilla Cyber Assurance Branch, Air Force Research Lab, Rome, NY, USA

Nnamdi Osia West Virginia University, Morgantown, WV, USA

Mersini Paschou Department of Computer Engineering & Informatics, School of Engineering, University of Patras, Rio Campus, Patras, Greece

Victor Pomponiu Agency for Science, Technology and Research, Singapore, Singapore

Praveen Rao Department of Computer Science and Electrical Engineering, University of Missouri-Kansas City, Kansas City, MO, USA

Evangelos Sakkopoulos Department of Computer Engineering & Informatics, School of Engineering, University of Patras, Rio Campus, Patras, Greece

Lisa Singh Georgetown University, Washington, DC, USA

Spyros Sioutas Department Informatics, Ionian University, Corfu, Greece

Efrosini Sourla Department of Computer Engineering & Informatics, School of Engineering, University of Patras, Rio Campus, Patras, Greece

Athanasios Tsakalidis Department of Computer Engineering & Informatics, School of Engineering, University of Patras, Rio Campus, Patras, Greece

Giannis Tzimas Department of Computer & Informatics Engineering, Technological Educational Institute of Western Greece, Patras, Greece

Emmanouil Viennas Department of Computer Engineering & Informatics, School of Engineering, University of Patras, Rio Campus, Patras, Greece

Yifang Wei Georgetown University, Washington, DC, USA

Wai L. Woo School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, UK

Xue Yang West Virginia University, Morgantown, USA

Konstantinos C. Zikidis Department of Aeronautical Sciences, Hellenic Air Force Academy, Dekelia Air Base, Greece

Part I
Surveillance of Human Features

Chapter 1
A Survey of Using Biometrics for Smart Visual Surveillance: Gait Recognition

Imed Bouchrika

Abstract In spite of the increasing concerns raised by privacy advocates against the intrusive deployment of large scale surveillance cameras, the research community has progressed with a remarkable pace into the area of smart visual surveillance. The automation for surveillance systems becomes an obligation to avoid human errors and ensure an efficient strategy for tackling crimes and preventing further terrorist attacks. In this research article, we survey the recent studies in computer vision on gait recognition for covert identification and its application for surveillance scenarios and forensic investigation. The integration of biometric technologies into surveillance systems is a major step milestone to improve the automation process in order to recognize criminal offenders and track them across different places. The suitability of gait biometrics for surveillance applications emerges from the fact that the walking pattern can be captured and perceived from a distance even with poor resolution video as opposed to other biometric modalities which their performance deteriorates in surveillance scenarios.

Keywords Biometrics · Gait recognition · Visual surveillance · Re-identification

Introduction

The deployment of surveillance systems has become ubiquitous in our society regardless of the increasing concerns on personal privacy and data confidentiality. Although privacy advocates fear large scale intrusive surveillance of private individuals, there is a non-debatable consensus that the security and safety of citizens are considered the most rudimentary requirements to be ensured and guaranteed against the sharp increase of the uncountable crimes and terrorist attacks that took place during the last decade across Europe, North Africa and the Middle East. Law enforcement agencies can make use of surveillance footage for the safety of

I. Bouchrika (✉)
Faculty of Science and Technology, University of Souk Ahras, Souk Ahras, Algeria
e-mail: imed@imed.ws

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_1

our neighborhood and crime prevention. The use of surveillance technology should without doubt assist to lessen the risks and number of crimes by serving as a deterrent. Biometric technologies can be a major milestone to improve the automation process of visual surveillance in order to recognize criminal offenders and track them across different places. Gait defined as the way we walk, is considered recently as a more suited modality for people recognition in surveillance and forensic scenarios. This is because it can be captured non-intrusively and covertly from a distance even with poor resolution imageries. Although, the distinctiveness of gait cannot compare with traditional biometrics, considerable body of research emerged recently asserting the potency and merits of gait biometrics in surveillance applications [28]. In this chapter, we survey the recent studies on the use of computer vision methods for gait biometrics covert identification and its application for surveillance scenarios and forensic investigation.

Smart Visual Surveillance

Visual surveillance systems work based on mounting cameras at remote locations in order to transmit video streams which are stored and monitored at real time. The placement of video cameras is installed at sensitive places and public areas with video being sent to the control center where video feeds are viewed by security staff on multiple screens. The deployment of visual surveillance systems plays a critical role in gathering vital intelligence for law enforcement officers to help them track and detect suspicious individuals who had or about to commit crimes. Visual surveillance systems span to a number of applications from observing people, vehicles, animals and military targets. Originally, visual surveillance technologies were designed for human agents to observe protected remote spaces concurrently as well as to record video data for further offline analysis. In fact, observing surveillance video is a labor-intensive activity specially when a large number of cameras need to be controlled simultaneously. Nowadays, with the sudden increase of surveillance cameras that have exceeded 5 million cameras within the United Kingdom alone, massive amount of video data is being transmitted and recorded. The issue of whether it is watched, reviewed or analyzed is still questionable due to the lack of human resources, time and sheer volume of data. Consequently, criminal incidents and suspicious events may be missed and not noticed in time to prevent a true crime from taking place. For installations with a large number of cameras, it is obviously impossible for one or even several operators to watch all the cameras. Even in the unrealistic scenario that there is one security staff per camera, an empirical study published by Security Oz [42] argued that after observing 12 min of continuous video monitoring, the operator will often miss afterwards up to 45% of screen activities. After 22 min of watching, up to 95% is overlooked. Therefore, if surveillance video is not monitored carefully and acted upon, there is obviously an increased security risk for individuals and private facilities.

Because of the impossibility to watch and screen all surveillance videos all the times, the installation of surveillance systems is rendered limited in terms of effectiveness and efficacy. There is a significant and increasing demand to address such limitations that posed as a great challenge for researchers and industrial practitioners to provide solutions and theories to meet the overwhelming global security needs. According to Liu et al. [37], there are over 6000 research articles published in the literature since 1971 covering the area of intelligent visual systems and analytics. The concept floated to appearance in early 1970s but emerged to the research community during the 80s. The research progressed at faster pace since 2000 due to development of computational resources and availability of cameras at cheap prices. Most of the published papers in visual surveillance fall into one of the three aspects including hardware, software and surveillance applications. New smart visual surveillance systems are currently being developed to take video sequences, pre-process the data using computer vision techniques and then analyze the data to learn and detect interesting events and objects. For instance, vehicle license plates can be automatically recognized or virtual fences can be setup around critical facilities so that warnings can be raised in the event of an unauthorized access or intrusion. Elliott [17] has recently defined an Intelligent Video System (IVS) as:

> any video surveillance solution that utilizes technology to automatically, without human intervention, process, manipulate under/or perform actions to or because of either the live or stored video images

Smart or intelligent visual surveillance system is designed to collect, index and store video data allowing security and law enforcement agents to monitor, analyze and search for events or objects of interest at real time or previous times. Further, along with the sheer volume of online and offline video footage coming from multiple cameras, it becomes a necessity to develop efficient methods for the retrieval and analysis of video data based on semantic content. Security and military applications have been the main thrust in developing smart surveillance tools. From small security outpost, shopping malls to large cities with thousands of citizens. Recently, as the number of cameras has increased immensely by both governmental and private agencies for the protection of their citizens, employees and properties. Smart visual applications analyze and filter out the massive amount of data recorded from multiple continuous video feeds ensuring only appropriate alerts are presented for law enforcement officers. The automation for surveillance systems becomes an obligation to avoid human errors and ensure an efficient strategy for tackling crimes and preventing further terrorist attacks. The process of automation can be fulfilled either in a centralized fashion such that all processing is done within a single place where all video streams are sent. Alternatively, processing can be performed in a distributed way where even cameras can embed certain intelligence for detecting events or recognizing people. The automation of intelligent video systems can include simple tasks from people counting, pedestrians detection to further complex and intricate tasks as behavior analysis [33] and people identification.

Biometric Technologies

Biometrics is defined as the automated process of extracting and deriving descriptive measurements based on either the behavioral or physiological characteristics of the human body. Such measurements should distinguish an individual uniquely among other subjects. The derived biometric data is compared against a database of records to either verify or recognize the identity of an individual. The word *biometrics* is a composite word of two parts from the Greek language: *bios* means life meanwhile *metrics* refers to the verb "to measure". Apart from the uniqueness condition, the biometric description should be universal and permanent. The universality factor implies that the biometric data can be acquired from all the population regardless of their gender, age, location or ethnicity. For the permanentness condition, it signifies that the biometric signature of an individual should stay the same throughout the different ages. Jain et al. [26] added further criteria that must be met for biometric systems including user acceptance, performance, unvulnerability and integration. As opposed to traditional identification or verification methods such as passports, passwords or pin numbers, biometrics cannot be transferred, forgotten or stolen and should be ideally obtained non-intrusively. Even though, people intuitively use some body characteristics such as face, gait or voice to recognize each other, it is proven a challenging task to extract and quantify such measurements into a biometric signature using computer vision methods.

The biometric measurements are based either on the physiological traits such as face, ear, fingerprint and DNA or based on the behavioral traits including gait, voice and signature. Figure 1.1 shows examples of both types of biometrics. Biometrics is now emerging in regular use being deployed in various applications such immigration border control, forensic systems, computer security and payment authentication. Biometric systems are sold mainly for the following purposes: physical access control, logging attendance and personal identification purposes. The choice of a specific biometric modality depends on the nature and requirements of the intended applications in addition to the availability of biometric features. Though in many cases, the combination of multiple biometric modalities may be needed to achieve the desired level of performance [46]. Fingerprints, iris and face are among the most popular physiological traits used successfully in commercial identification systems with fingerprint capturing over 50% of the market share [27]. The main reason for such popularity is the availability of large legacy databases which have been collected by law enforcement agencies all over the world [28]. There is no doubt that biometrics has a clear-cut benefits over the use of passwords or identification cards, there are still limitations that biometrics are vulnerable to such as spoofing, linkability attacks, evasion and database alteration [18, 28].

Biometric systems are setup to work either in identification or verification mode. For identification, a one-to-many matching process is conducted for newly acquired biometric data against all people already enrolled in the database in order to infer the identity of the subject whose matched biometric data exhibits the highest similarity value. For the case when the identification system is forced to infer the identity of

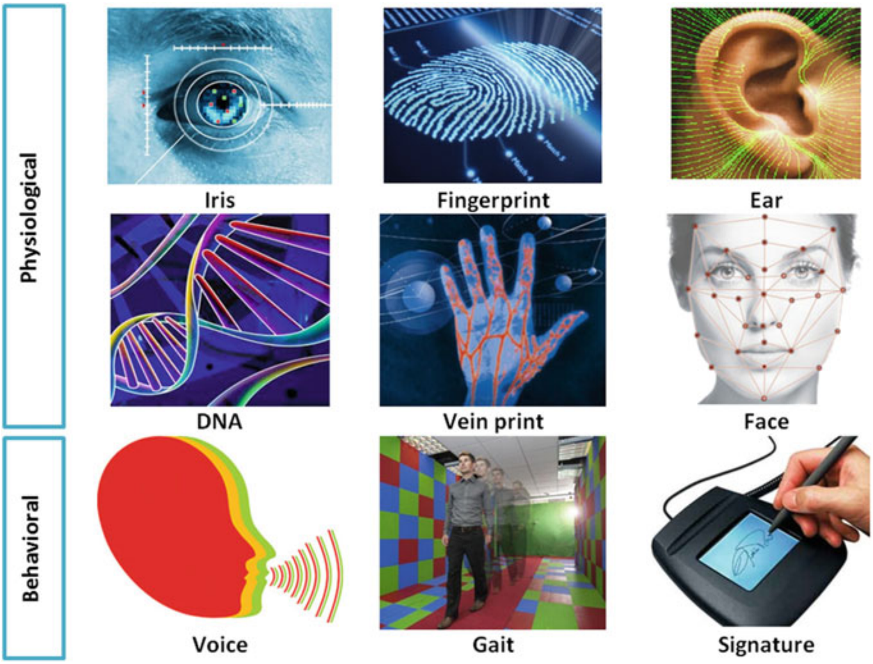


Fig. 1.1 Various types of biometric modalities

the unknown person from the list of people enrolled in the database, it is referred to as a *closed-set* biometric system. In contrast to the *open-set* biometric system when there can be an option to have a *reject* or *no match* response. For the verification mode, the system conducts a one-to-one match for the person's signature against a pre-recorded verified signature in the database to confirm the claimed identity. The identity claim is confirmed if the computed similarity score is above a preset threshold. Figure 1.2 shows an overview for a biometric system with the different components and steps involved from extracting the measurements to the matching process to infer the person identity.

For the history of biometrics, people have been recognizing each other based on voice, face or walking style for thousands of years. However, the first systematic basis for people identification dates back to 1858 when William Herschel recorded the handprint of each employee on the back of a contract whilst working for the civil service of India [7]. This was used as a way to distinguish staff from each other on payday. It was considered the first systematic capture of hand and fingerprints that was used primarily for identification purposes. For the use of biometrics in forensic analysis, Alphonse Bertillon who was a French police officer was the pioneer to use the first biometric evidence into the judicial system presented as anthropometric measurements of the human body to counter against repeat offenders who could easily change or fake their names and personal details. In 1879, he introduced the

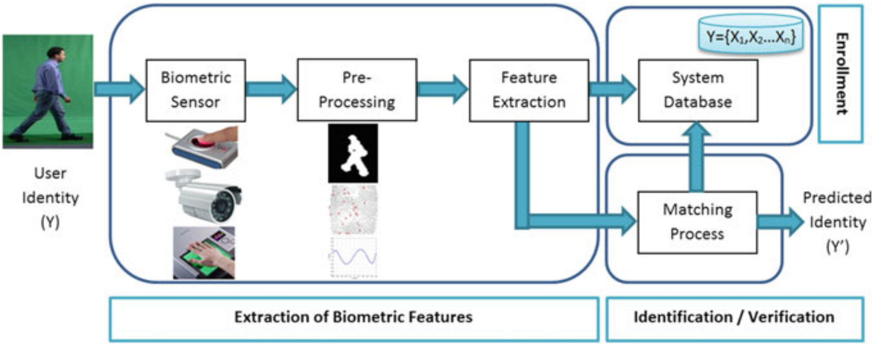


Fig. 1.2 Overview of a biometric system

bertillonage system such that every individual is identified through detailed records taken from their body anthropometric data and physical descriptions as they are impossible to spoof or change them. The system was adopted by the police authorities throughout the world. In 1903, the Bertillon system was scrutinized on the basis that anthropometric measurements are not adequate to differentiate reliably between people. This was fueled by the case of identical twins who have almost the same measurements using the Bertillon system. The first scientific research paper on the use of fingerprints for identification was published in 1880 by Henry Faults in Nature [28]. During 1890s, Sir Francis Galton and Edward Henry proposed separately classification systems for people recognition based on their fingerprints taken from all the ten fingers [2]. The characteristics that Galton described to identify people are still used today. The Galton system was adopted by New York state prison where fingerprints were utilized for the identification of apprehended inmates.

Gait Recognition

In spite of the fact that physiological biometric systems enjoy the merit of reliable identification rates, their use in surveillance systems is never considered due to the limitation of extracting them unintrusively. Interestingly, gait biometrics is reported by recent studies to be suitable for surveillance scenarios because it can be captured from a distance even with poor resolution imageries in contrast to other biometric modalities which require high resolution images. Moreover, the main potency of gait biometrics is the non-invasiveness property and therefore the person is not obliged to cooperate or interact with the acquisition hardware. The invasiveness property sets the identification procedure via gait ideal for cases where direct contact with the offender is not an option which is the case for all criminal cases. Moreover, a biometric signature constructed from the gait rhythmic motion pattern is considered the only likely identification method suitable for covert surveillance and reliably not prone to spoofing attacks and signature forgery. Consistently, many recent research studies concluded that gait recognition is more suitable for forensic science as other identification traits that can be related to the crime scene can be wiped out or concealed in contrast to the gait motion as the mobility of the person is a must for them to walk away from the crime scene. Recently, Lucas et al. [39] reported that a combination of eight anatomical measurements from the human body is enough to attain a probability to the order of 10^{-20} for a finding duplicate signature comparing such results to fingerprint analysis. Interestingly, one of the murder cases that attracted the media attention in the United Kingdom where a child was kidnapped and murdered, it was impossible for the investigating team to infer the identity of the killers from poor resolution surveillance video footage. The only inspiring method that could be used to reveal their identities in such delicate situation was the gait pattern as suggested by the research team from the University of Southampton [43]. The notion that people can be recognized by the way they walk has gained an increasing popularity and produced impacts on public policy and forensic practice by its take up by researchers at the Serious Organized Crime Agency.

In 1964, Murray et al. [41] performed the early experiments describing the standard gait pattern for normal walking people aimed at studying the gait pattern for pathologically abnormal patients. The medical investigations were conducted on sixty individuals aged between 20 and 65 years old. Each subject was instructed to walk for a repeated number of trials. For the collection of gait data, special markers were attached on every person. Murray et al. [41] suggested that human gait consists of 24 different components which make the gait pattern distinctive for every subject if all gait movements are considered. It was reported that the motion patterns of the pelvic and thorax regions are highly variable from one person to another. Furthermore, the study reported that the ankle rotation, pelvic motion and spatial displacements of the trunk embed the subject individuality due to their consistency at different experiments. In 1977, Cutting et al. [14] published a paper confirming the possibility of recognizing people by the way they walk via observing Moving Lights Displays (MLD) mounted on the joints positions. An MLD shown in Fig. 1.3 is a two-dimensional video of a collection of bright dots attached to the human body taken against a dark background where only the bright dots are visible in the scene. Different observers are asked to see the actors performing various activities. Based

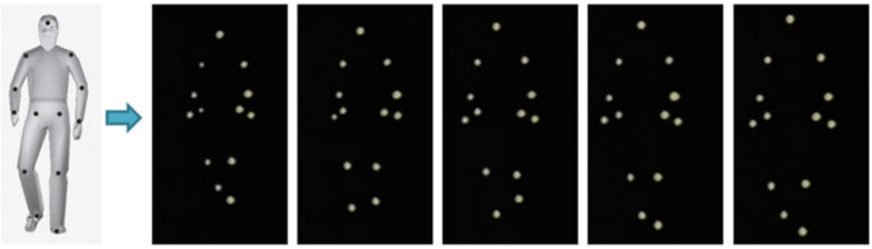


Fig. 1.3 Medical studies

on these experiments observers can recognize different types of human motion such as walking, jumping, dancing and so on [29]. Moreover, the observer can make a judgment about the gender of the performer [14], and even further identify the person if they are already familiar with their gait [14, 19]. Cutting argued that the recognition is purely based on dynamic gait features as opposed to previous studies which were confounded by familiarity cues, size, shape or other non-gait sources of information. Although, there is a wealth of gait studies in the literature aimed for medical use with only a few referring to the uniqueness nature of the walking pattern, none is concerned with the automated use of the gait pattern for biometrics. The gait measurements and results from the medical literature introduced by Cutting, Murray and Johansson are to be of benefit for the development of automated gait biometric systems using computer vision methods though the extraction of the gait pattern is proven complex.

Gait Recognition Methods

Vision-based system for people recognition via the way they walk, is designed to extract gait features without the need to use special sensors or reflective markers to assist the extraction process. In fact, all that is required is a video camera to stream images for vision-based software for processing. Marker-less motion capture systems are suited for applications where mounting sensors or markers on the subject is not an option as the case of outdoor surveillance applications. Typically, gait biometric system consists of two main components: (i) a hardware platform dedicated for data acquisition. This can be a single CCTV camera or distributed cameras network. (ii) A software platform for video processing and identification. The architecture of the software tier for gait biometrics is composed broadly of three main stages: (i) *detection and tracking of the pedestrian*: intra-camera tracking is performed to establish the correspondence of the same person across consecutive frames. (ii) *Feature extraction*: in order to estimate a set of measurements either related to the configuration of the whole body or the configuration of the different body parts in a given scene and tracking them over a sequence of frames. (iii) *Classification stage*: which involves matching a test sequence with an unknown label against a group of labelled references considered as the gallery dataset. Figure 1.4 shows the flow diagram for gait identification outlining the different subsystems involved in the process of an automated people recognition.

Much of the interest in the field of human gait analysis was limited to physical therapy, orthopedics and rehabilitation practitioners for the diagnosis and treatment of patients with walking abnormalities. As gait has recently emerged as an attractive biometric, gait analysis has become a challenging computer vision problem. Although, the distinctiveness of gait features cannot compare with traditional biometrics, it has proven to be a potential alternative for surveillance scenarios [28]. Many research studies have aimed to develop a system capable of overcoming the difficulties imposed by the extraction and tracking of biometric gait features. Various

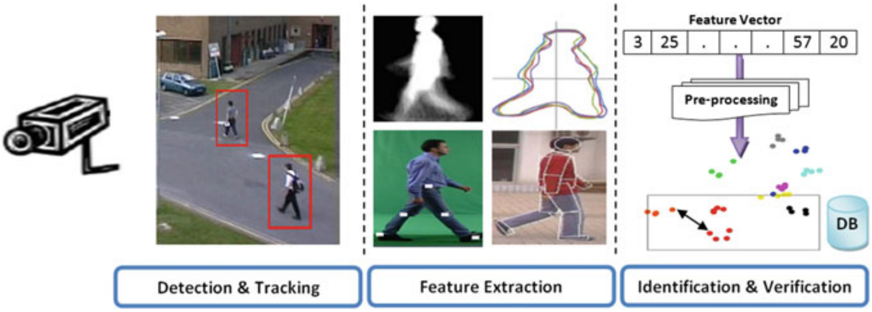


Fig. 1.4 Overview of gait biometric system

methods were surveyed in [43] and [56]. Based on the procedure for extracting gait features, gait recognition methods can be divided into two main categories which are model-based and appearance-based (model-free) approaches.

Appearance-Based Approaches

Appearance-based or model-free approaches for gait recognition do not need a prior knowledge about the gait model. Instead, features are extracted from the whole body without the need to explicitly extract the body parts. The majority of appearance approaches depends on data derived from silhouettes which are obtained via background subtraction. Appearance-based Method relies pivotally on statistical methods to reduce or optimize the dimensionality of feature space using methods such as Principal Component Analysis. In addition, advanced machine learning methods are usually applied as multi-class support vector machine and neural networks. Contentiously, recent investigations by Veres et al. [51] reported that most of the discriminative features for appearance-based approaches are extracted from static components of the top part of the human body whilst the dynamic components generated from the swinging of the legs are ignored as the least important information.

Silhouette-Based Methods

Silhouette-based methods work by separating walking people from the background. The simplest baseline method is to compute the similarity score between two synchronized silhouette sequences [47]. The *Gait Energy Image (GEI)* is another basic silhouette-based representation introduced by Han and Bhanu [21] in which gait signature is constructed through taking the average of silhouettes for one gait cycle. Experimental results confirmed that higher recognition rates can be attained to reach 94.24% for a dataset of 3,141 subjects [25]. However, such method performs poorly when changing the appearance. The *Motion Silhouette Image* (MSI) is a similar representation to GEI proposed by Lam et al. [34] where each pixel intensity is computed as a function of the temporal history of motion for the corresponding pixels across a complete gait cycle. Experiments conducted on the large SOTON gait dataset showed that 87% can be achieved. *Gait Entropy Image (GenI)* is a silhouette-based representation introduced by Bashir et al. [4] which is computed by calculating the Shannon entropy for each pixel achieving a correct classification rate of 99.1% on dataset of 116 subjects. The Shannon entropy estimates the uncertainty value associated with a random variable. Other similar representations include Motion Energy Image, Gait History Image, Frieze Patterns and Chrono-Gait Image.

For using the gait symmetric property, Hayfron-Acquah et al. [22] introduced a method for constructing a gait signature based on analysing the symmetry of human motion. The symmetry map is produced via applying the Sobel operator on the gait silhouettes followed by the Generalized Symmetry Operator. The symmetry map was evaluated on a dataset containing 28 people using the k-NN classifier, a high recognition rate of 96.4% was attained for the value of $k = 3$. There is a recent tendency to use model-free depth based representation using 3D sensors (Fig. 1.5). Sivapalan et al. [48] proposed the *Gait Energy Volume* descriptor (GEV) by extending the Gait Energy Image into 3D. The implementation for GEV was evaluated on the CMU MoBo database confirming that improvements can be attained over the 2D GEI version as well as fused multi-view GEI variant. Recently, there is a trend of employing Deep Learning and Neural Networks using Silhouette-based descriptors to account for the issue of view-invariance. Li et al. [36] proposed a gait representation called *DeepGait* using deep convolution features. Meanwhile, Wu et al. [55] utilized deep Convolution Neural Networks (CNNs) for gait recognition using the OU-ISIR gait dataset with a reported success rate of 94% for a population of 4,007 subjects. Zeng et al. [59] described a silhouette-based approach for view-invaraint gait biometrics using deterministic learning theory. Radial basis function (RBF) neural networks are used to locally approximate the gait dynamics from different view angles.

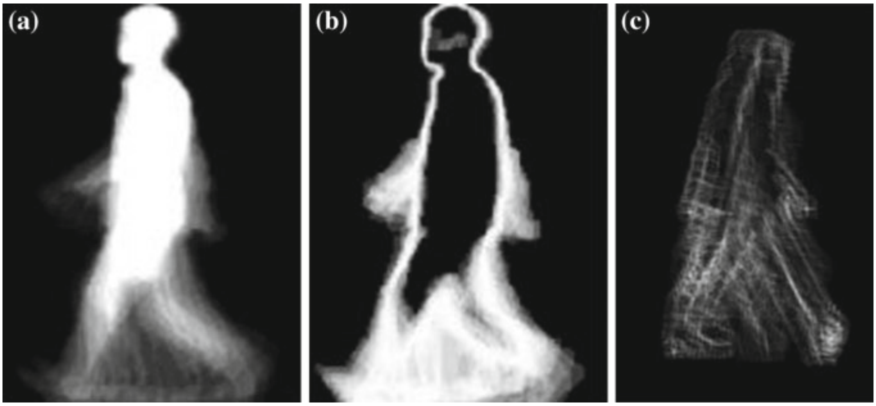


Fig. 1.5 Silhouette-based methods for gait recognition: **a** use gait energy image [21] **b** gait entropy image [4] **c** symmerty map [22]

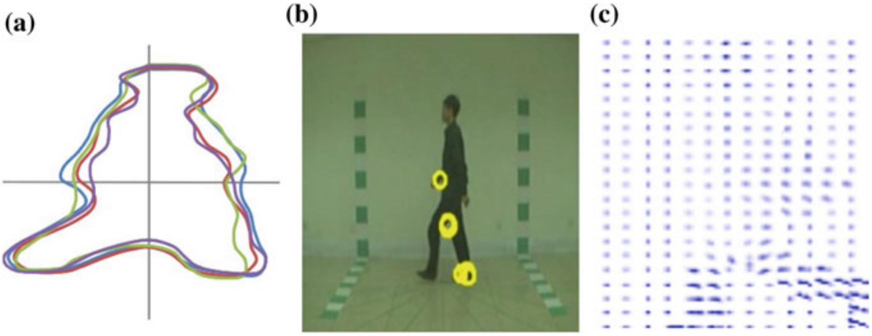


Fig. 1.6 Non-silhouette appearance-based methods for gait recognition: **a** procruste shape analysis [12] **b** STIP descriptors **c** optical flow

Non-silhouette Methods

As the accuracy of silhouette-based methods depend on the background segmentation algorithm which is not reliable for the case of real surveillance footage in addition to the sensitivity issue to varied appearances, a number of appearance-based methods have emerged recently that use instead interest-point descriptors. Kusakunniran et al. [32] proposed a framework to construct gait signature without the need to extract silhouettes. Features are extracted in both spatial and temporal domains using Space-Time Interest Points (STIPs) by considering large variations along both spatial and temporal directions at a local level. Bashir et al. [4] used the dense optical flow field computed using the method in [11] for each frame of the whole gait cycle to extract four different types of motion descriptors (Motion Intensity Image and four Motion Direction Images) based on the horizontal and vertical optical flow components and magnitude, their proposed experiments on the CASIA and SOTON gait datasets with the clothing and bag carrying covariates outperform previous reported studies. Hu et al. [23] described an incremental learning framework using optical flow for gait recognition. The Local binary pattern (LBP) operator is utilized to encode the textural information of optical flow (Fig. 1.6).

Model-Based Approaches

For the model-based approach, a prior model is established to match real images to this predefined model, and thereby extracting the corresponding gait features once the best match is obtained. Usually, each frame containing a walking subject is fitted to a prior temporal or spatial model to explicitly extract gait features such as stride distance, angular measurements, joints trajectories or anthropometric measurements. Although model-based approaches tend to be complex requiring high computational cost, these approaches are the most popular for human motion analysis due to their advantages [56]. The main strength of model-based techniques is the ability to extract detailed and accurate gait motion data with better handling of occlusion, self-occlusion and other appearance factors as scaling and rotation. The model can be either a 2 or 3-dimensional structural model, motion model or a combined model.

2D Structural Models

The structural model describes the topology of the human body parts as head, torso, hip, knee and ankle by measurements such as the length, width and positions. This model can be made up of primitive shapes based on matching against low-level features as edges. The stick and volumetric models are the most commonly used structural-based methods. Akita et al. [1] proposed a model consisting of six segments comprising of two arms, two legs, the torso and the head. Guo et al. [20] represented the human body structure by a stick figure model which had ten articulated sticks connected with six joints. Rohr et al. [45] proposed a volumetric model for the analysis of human motion using 14 elliptical cylinders to model the human body. Karaulova et al. [30] used the stick figure to build a hierarchical model of human dynamics represented using Hidden Markov Models. For the deployment of structural model-based methods for gait recognition, Niyogi et al. [44] was perhaps the pioneer in 1994 to use a model-based method for gait recognition. Gait signature is derived from the spatio-temporal pattern of a walking subject using a five stick model. Using a database of 26 sequences containing 5 different subjects, a promising classification rate of 80% was achieved. The recent trend of model-based approaches has shifted towards combining different cues including motion and 3D data in order to construct models able to handle the extraction of gait features (Fig. 1.7).

Motion Models

The motion model describes the kinematics or dynamics of the human body or its different parts throughout time. Motion models employ a number of constraints that aid the extraction process as the maximum range of the swinging for the low limbs. Cunado et al. [13] was the first to introduce a motion model using the Velocity Hough Transform to extract the hip angular motion via modeling human gait as a moving

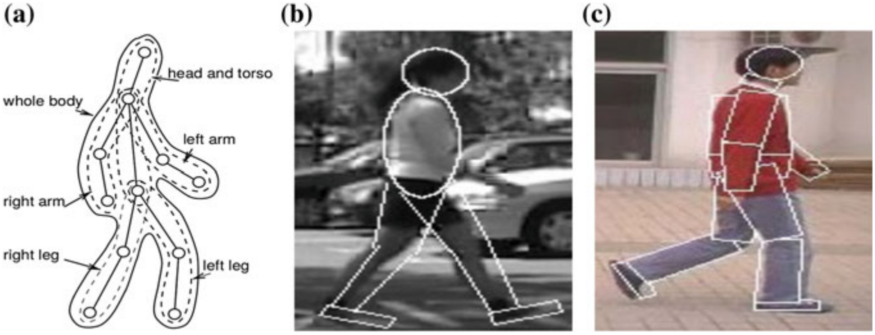


Fig. 1.7 Model-based approaches: **a** Karaulova [30]. **b** Wagg [52]. **c** Wang [53]

pendulum. The gait signature is derived as the phase-weighted magnitudes of the Fourier components. A recognition rate of 90% was achieved using the derived signature on a dataset containing 10 subjects. Yam et al. [57] modeled the human gait as a dynamic coupled oscillator which was utilized to extract the hip and knee angular motion via evidence gathering. The method was tested on a small dataset of 20 walking and running people, achieving a recognition rate of 91% based on gait signature derived from the Fourier analysis of the angular motion. Wagg et al. [52] proposed a new model-based method for gait recognition based on the biomechanical analysis of walking people. Mean model templates are adopted to fit individual subjects. Both the anatomical knowledge of human body and hierarchy of shapes are used to reduce the computational costs. The gait features vector is weighted using statistical analysis methods to measure the discriminatory potency of each feature. On the evaluation of this method, a correct classification rate of 95% is reported on a large database of 2,163 video sequences containing 115 different subjects (Fig. 1.7). Bouchrika et al. [9, 10] proposed a motion-based model for the extraction of the joints via the use a parametric representation of the elliptic Fourier descriptors describing the spatial displacements. The evaluation was carried out on a limited dataset containing 20 people from the Southampton dataset reporting a correct classification rate of 92% using the k-NN classifier.

3D Models

As most of the model-based methods are exploiting 2D images, there are recent work aimed for introducing 3-dimensional models for the extraction of gait features. 3D approaches for gait identification are known for their robustness of invariance to different viewpoints. Though, it is always a difficult task to acquire 3D data in surveillance data in addition to the high computational and financial costs involved. Recent studies on using 3D models include the work of Ariyanto et al. [3] who introduced a 3D approach using a marionette and mass-spring model to gait biometrics with 3D voxel gait data. The articulated human body is modeled using the stick-figure which emulates the marionettes' motion and joint structure. The stick-figure is composed of 11 nodes corresponding to the human joints including the head. Zhao [60] employed local optimization algorithms in order set up the 3D model from video sequences captured from multiple cameras. Features for the classification are derived from body segments including static parameters in addition to dynamic features for the motion trajectories of the limbs. Recently, Tang et al. [50] proposed an approach for gait partial similarity matching which is based on the assumption that 3-dimensional objects share similar view surfaces across different views. 3D parametric models for the human body are morphed by pose and shape deformation from a template model using 2-dimensional silhouette data. López-Fernández et al. [38] used 3D angular data to create a view-invariant gait biometric signature for people walking on unconstrained paths. Support vector machine is used for the classification process on two datasets containing 62 subjects in total. Kastaniotis et al. [31] used the Microsoft Kinect sensor to recognize people via a pose estimation process to extract the skeleton of a walking person. Based on a publicly available dataset of 30 people, a high recognition rate of 93.3% is reported.

Areas of Applications

Re-Identification

Due to the unprecedented surge for the number of crimes and terror attacks, the deployment of surveillance technology has become ubiquitous on our modern society. It is no wonder that many military, corporate and government agencies devoted a large amount of funding to research institutions to advance research for the deployment of biometric technologies within their surveillance systems to ease the management and monitoring for ensuring the safely of their citizens or assets. Due to the limitation inherited from using a single camera which can only cover a limited field of view, the deployment of interconnected set of cameras can becomes more prevalent in sensitive and crowded areas including shopping malls and airports. Assuming that pedestrians tracking within a single camera is adequately accurate and reliable, the re-identification issue is reduced thereafter to associating subjects of the same identity across different cameras. People re-identification is the task of tracking people regardless of their identity across different cameras [16]. The process to trace and follow the whereabouts of pedestrians and possibly producing a semantic description of their behaviour from visited places would be a major breakthrough to automated surveillance for law enforcement officers.

As tracking people within a single field of view can be performed reliably and robustly, it is a difficult task to have the tracking done within a network of cameras when it comes to solve the association problem for the same subjects seen at different places, different times from different cameras without overlapping views. The complexity for re-identification is mainly exacerbated from the variation of people appearances across different cameras [6]. This is besides other challenging factors including image quality, crowded scenes and occlusion. The articulated property of the human body is another prime reason for the appearance to subtly change continuously. Surveillance systems consisting of interconnected cameras cover large spatial areas with non-overlapping views to yield enhanced coverage. Therefore, the complexity of tracking people would proportionally increase with size of the area and number of cameras rending the process of re-identification and extremely resource intensive task temporally and spatially. There are a number of databases being setup and made publicly available for researchers to tackle the problem of people re-identification for surveillance applications. Bedagkar-Gala [6] published recently a survey on the different methods for people re-identification, performance metrics and datasets. The author classifies the techniques for people re-identification into three main categories: (i) Descriptor learning which is based on deriving the most discriminative characteristics on the fly or instead a learning phase is setup in order to derive a descriptive dictionary of features that better represent distinctively a person's appearance using bag-of-features approach. (ii) Distance metric learning aims to maximize the matching score between different identities regardless of the choice for appearance cues or color calibration using different types of features. (iii) Color Calibration approaches is based on modeling the color relationship across different cameras and update such model regularly for each camera. The color brightness transfer function (BTF) is employed to establish the association between the same people across different camera viewpoints.

Gait is an emergent biometrics which has attracted unprecedented interest from the research community. It enjoys the potency to overcome most of the drawbacks that other biometric modalities suffer from. Gait biometrics is deemed suitable for re-identification applications partly due to the fact that the gait motion can be recorded and extracted from a distance besides the non-invasive and less-intrusive property. For the process of markerless extraction of gait features from surveillance videos, a Haar-like template matching procedure is presented by Bouchrika [10] to locate the spatial positions of the lower limbs for a single walking person. The method is not dependent on foreground segmentation for the extraction of gait features. This is mainly because it is resource intensive and expensive process to perform background subtraction for real-time scenarios due to the task of updating the background model. The motion models are taken from the medical data reflecting the angular motion for the hip and knee within one full gait cycle. For the initial phase, the described approach builds a motion map from the change detection of the inter-frame differences. The drawback of depending on frame differencing is that the camera has to be stationary. Moving pixels belonging to moving people across successive frames are extracted with the emphasis to extract clear contour data. Bedagkar-Gala [5] combined the use of colors with gait features including Gait Energy Image and Frame Difference Energy Images in order to track 40 people across 9 different cameras taken from the SAIVT SoftBio dataset. Wei et al. [54] proposed a Swiss-system based cascade ranking model for re-identification of people in surveillance using five gait silhouette-based descriptors. The Swiss-system uses multi-feature ensemble learning where a series of rankers are applied to every pair of matches.

Forensic Analysis

Forensic gait analysis has been recently applied in investigations at numerous criminal cases as law enforcement officers have no option to identify the perpetrator using well-established methods as facial recognition or fingerprints. This is partly due to the fact that key biometric features such as the perpetrator's face can be obscured or veiled and the CCTV footage is deemed unusable for direct recognition whilst the perpetrators are usually filmed at a distance walking or running away from the crime scene. Gait experienced specialists are consulted to assist with the identification process of an unknown person by their walking pattern through a qualitative or quantitative matching process. This would involve examining the unique and distinctive gait and posture features of an individual. Subsequently, a statement is written expressing an opinion or experimental results that can be used in a court of law. Meanwhile, the practice of forensic podiatry involves examining the human footprint, footwear and also the gait pattern using clinical podiatric knowledge [15]. However, gait analysis performed by a podiatrist involves the recognition and comparison of nominal and some ordinal data without quantitative analysis using numerical forms of data [15]. Because of the rising profile of gait biometrics and forensic podiatry, gait is used numerously as a form of evidence in criminal prosecutions with the inauguration of the American Society of Forensic Podiatry in 2003. Recently, Iwama et al. [24] developed a software with a graphical user interface in order to assist non-specialists to match video sequences based on gait analysis. For the methods used for forensic gait analysis, Bouchrika et al. [8] classifies them into two major categories which are: *Descriptive-based* or *Metric-Based* approaches.

An incident of a bank robbery in 2004 was handled by the Unit of Forensic Anthropology at the University of Copenhagen [40]. The police observed that the perpetrator has a special gait pattern with a need to consult gait practitioners to assist with the investigation. The police were instructed to have a covert recording of the suspect walking pattern within the same angle as the surveillance recordings for consistent comparison. The gait analysis revealed that there are several matches between the perpetrator and the suspect as an outward rotated feed and inverted left ankle during the stance phase. Further posture analysis using photogrammetry showed that there is a resemblance between the two recordings including a restless stance and anteriour head positioning. There were some incongruities observed during the analysis including wider stance and the trunk is slightly leaned forward with an elevated shoulders. This is suspected to be related by the anxiety when committing a crime [35]. Based on the conducted analysis, a statement was given to the police regarding the identity however such methods are argued that they do not constitute the same level of confidence as well-established methods such as fingerprints. The findings were subsequently presented in court and the suspect was convicted of robbery whilst the court stressed that gait analysis is a valuable tool [35]. In a similar case handled by the Unit of Forensic Anthropology, a bank robbery was committed by two masked people wearing white clothing. The bank was equipped with several cameras capturing most of the indoor area. One of the camera showed one of the perpetrator walking rapidly from the entrance. The frame rate was low which left only few useful images showing the perpetrator gait. Based on experimental results showing the most discriminatory potency for the joints angles, Yang et al. [58] argued about the possibility of identification based on certain instances of the gait cycle using the observed angular swinging of the legs. Figure 1.8 shows the two discussed cases of the bank robberies handled by the forensic unit.

In a recent case handled by the Metropolitan Police of London [8], a number of crimes include physical assaults and burglary against pedestrians walking on a short pathway near a subway in one of the London suburb. The same crime was reported to occur numerous times in the same fashion and at the same place. The police officers strongly suspected it was carried out by the same members of an organized gang of youngsters aged between 17 and 20 years old. There are a number of CCTV cameras in operation at the crime scene. Two of them are pointing towards the entrances of the subway as shown in Fig. 1.9. Two other cameras are set to record both views of the walking pass next the subway. The police provided a set of videos in order to deploy gait analysis to find further information that would assist them in their investigation. CCTV footage from all cameras for the crime scene at two different

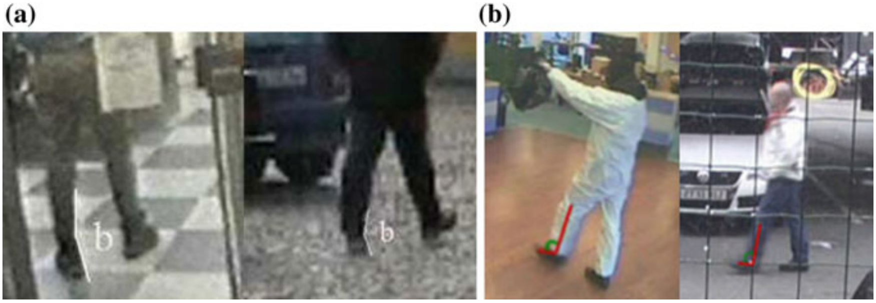


Fig. 1.8 Forensic gait analysis using descriptive-based methods [35, 58]: **a** 2004 **b** 2012

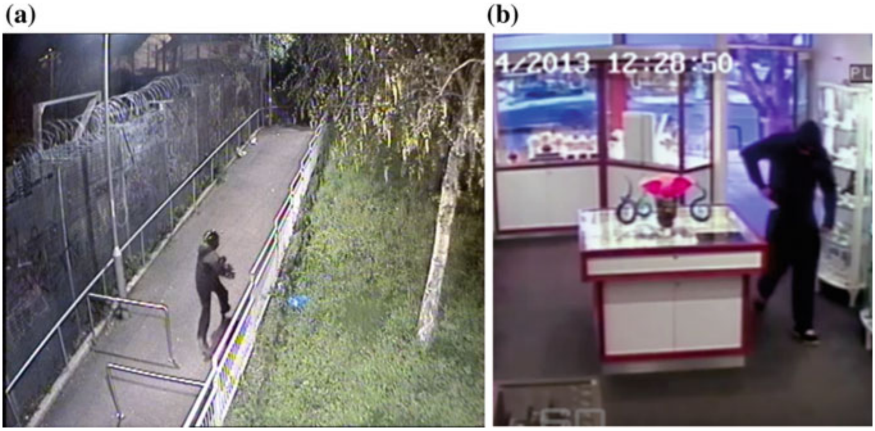


Fig. 1.9 Forensic gait analysis using metric-based methods: **a** 2011, UK case [8] **b** 2013, Australian case [49]

days was made available to the Image Processing Research group at the University of Southampton. The police provided another video of a suspect member of the gang being recorded whilst was being held at the police custody. The video was recorded at a frame rate of 2 frames per second and a resolution of 720 × 576 pixels. In one of the videos that was recorded on 4th April 2008, two members of the gang wore helmets to cover their faces and drove a scooter motorbike. A female pedestrian came walking through the subway where they followed her from behind on the walking path. When she entered the subway, one of them walked and snatched her bag violently using physical assault and even dragging her down on the ground. Afterwards they left away on a scooter. In a different CCTV footage recorded on the following day, the same crime was carried out with apparently the same looking perpetrators riding a scooter motorbike seen snatching a bag of another woman. The Instantaneous Posture Matching (IPM) [8] is proposed to conduct the forensic analysis which is based on computing the normalized distances for the joints positions between two video sequences on a specific window of frames. To access the measurement confidence, empirical study is conducted to explore the intra- and inter-match scores produced by Instantaneous Posture Matching on a larger gait dataset. Sun et al. [49] suggested the potential use of gait acceleration for crime analysis via the detection of heel strikes.

Conclusion

Because of the unprecedented number of crimes and terror attacks as well as the vital need to provide safer environment, a surge of concerns has emerged in many countries to ensure the safety of their citizens via the use of advanced surveillance technology. It is no wonder that many cities have deployed a number of biometric technologies within their surveillance systems to ease the management and monitoring from a large number of cameras. Biometrics can be of benefits not only for identify recognition, but it can play a vital role to enhance the automation process for surveillance systems. Gait is an emergent biometrics which has attracted unprecedented interest from the research community. In contrast to other biometric modalities which requires high resolution images, gait is suitable for covert recognition in surveillance scenarios mainly to the non-invasiveness property and therefore the person is not obliged to cooperate or interact with the acquisition hardware.

References

1. Akita K (1984) Image sequence analysis of real world human motion. Pattern Recognit 17(1):73–83
2. Arbab-Zavar B, Xingjie W, Bustard J, Nixon MS, Li CT (2015) On forensic use of biometrics. In: Handbook of digital forensics of multimedia data and devices
3. Ariyanto G, Nixon MS (2012) Marionette mass-spring model for 3d gait biometrics. In: 5th international conference on biometrics. IEEE, pp 354–359
4. Bashir K, Xiang T, Gong S (2009) Gait recognition using gait entropy image. In: 3rd international conference on crime detection and prevention, pp 1–6
5. Bedagkar-Gala A, Shah SK (2014a) Gait-assisted person re-identification in wide area surveillance. In: Asian conference on computer vision. Springer, pp 633–649
6. Bedagkar-Gala A, Shah SK (2014b) A survey of approaches and trends in person re-identification. Image Vis Comput 32(4):270–286
7. Berry J, Stoney DA (2001) The history and development of fingerprinting. Adv Fingerprint Technol 2:13–52
8. Bouchrika I (2017) Evidence evaluation of gait biometrics for forensic investigation. In: Multimedia forensics and security. Springer, pp 307–326
9. Bouchrika I, Nixon MS (2006) Markerless feature extraction for gait analysis. In: Proceedings of IEEE SMC chapter conference on advanced in cybernetic systems, pp 55–60
10. Bouchrika I, Carter JN, Nixon MS (2016) Towards automated visual surveillance using gait for identity recognition and tracking across multiple non-intersecting cameras. Multimed Tools Appl 75(2):1201–1221
11. Brox T, Bruhn A, Papenberg N, Weickert J (2004) High accuracy optical flow estimation based on a theory for warping. In: European conference on computer vision. Springer, pp 25–36

12. Choudhury SD, Tjahjadi T (2012) Silhouette-based gait recognition using procrustes shape analysis and elliptic fourier descriptors. Pattern Recognit 45(9):3414–3426
13. Cunado D, Nixon MS, Carter JN (2003) Automatic extraction and description of human gait models for recognition purposes. Comput Vis Image Underst 90(1):1–41
14. Cutting JE, Kozlowski LT (1977) Recognizing friends by their walk: gait perception without familiarity cues. Bulletin Psychon Soc 9(5):353–356
15. DiMaggio JA, Vernon W (2011) Forensic podiatry principles and human identification. In: Forensic podiatry. Springer, pp 13–24
16. Doretto G, Sebastian T, Tu P, Rittscher J (2011) Appearance-based person reidentification in camera networks: problem overview and current approaches. J Amb Intel Human Comput 2(2):127–151
17. Elliott D (2010) Intelligent video solution: a definition. Security 47(6)
18. Evans N, Marcel S, Ross A, Teoh ABJ (2015) Biometrics security and privacy protection. IEEE Signal Process Mag 32(5):17–18
19. Goddard NH (1992) the perception of articulated motion: recognizing moving light displays. PhD thesis, University of Rochester
20. Guo Y, Xu G, Tsuji S (1994) Understanding human motion patterns. In: Pattern recognition, conference B: computer vision & image processing, proceedings of the 12th IAPR international conference on 2
21. Han J, Bhanu B (2006) Individual recognition using gait energy image. IEEE Trans Pattern Anal Mach Intel 28(2):316–322
22. Hayfron-Acquah JB, Nixon MS, Carter JN (2003) Automatic gait recognition by symmetry analysis. Pattern Recognit Lett 24(13):2175–2183
23. Hu M, Wang Y, Zhang Z, Zhang D, Little JJ (2013) Incremental learning for video-based gait recognition with lbp flow. IEEE Trans Cybern 43(1):77–89
24. Iwama H, Muramatsu D, Makihara Y, Yagi Y (2012a) Gait-based person-verification system for forensics. In: IEEE fifth international conference on biometrics: theory, applications and systems (BTAS), pp 113–120
25. Iwama H, Okumura M, Makihara Y, Yagi Y (2012b) The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Trans Inf Forensics Secur 7(5):1511–1521
26. Jain A, Ross AA, Nandakumar K (2011) Introduction to biometrics. Springer Science & Business Media
27. Jain AK, Kumar A (2012) Biometric recognition: an overview. In: Second generation biometrics: the ethical, legal and social context, pp 49–79
28. Jain AK, Nandakumar K, Ross A (2016) 50 years of biometric research: accomplishments, challenges, and opportunities. Patt Recognit Lett
29. Johansson G (1973) Visual perception of biological motion and a model for its analysis. Perception and psychophysics 14:201–211
30. Karaulova IA, Hall PM, Marshall AD (2000) A hierarchical model of dynamics for tracking people with a single video camera. In: Proceedings of the 11th british machine vision conference 1:352–361
31. Kastaniotis D, Theodorakopoulos I, Economou G, Fotopoulos S (2016) Gait based recognition via fusing information from Euclidean and Riemannian manifolds. Pattern Recognit Lett 84:245–251
32. Kusakunniran W (2014) Recognizing gaits on spatio-temporal feature domain. IEEE Trans Inf Forensics Secur 9(9):1416–1423
33. Ladjailia A, Bouchrika I, Merouani HF, Harrati N (2015) On the use of local motion information for human action recognition via feature selection. In: 4th international conference on electrical engineering (ICEE), 2015. IEEE, pp 1–4
34. Lam TH, Lee RS (2006) A new representation for human gait recognition: motion silhouettes image (MSI). In: International conference on biometrics. Springer, pp 612–618
35. Larsen PK, Simonsen EB, Lynnerup N (2008) Gait analysis in forensic medicine. J Forensic Sci 53(5):1149–1153

36. Li C, Min X, Sun S, Lin W, Tang Z (2017) Deepgait: a learning deep convolutional representation for view-invariant gait recognition using joint bayesian. Appl Sci 7(3):210
37. Liu H, Chen S, Kubota N (2013) Intelligent video systems and analytics: a survey. IEEE Trans Ind Inform 9(3):1222–1233
38. López-Fernández D, Madrid-Cuevas FJ, Carmona-Poyato A, Muñoz-Salinas R, Medina-Carnicer R (2016) A new approach for multi-view gait recognition on unconstrained paths. J Vis Commun Image Represent 38:396–406
39. Lucas T, Henneberg M (2015) Comparing the face to the body, which is better for identification? Int J Legal Med 1–8
40. Lynnerup N, Vedel J (2005) Person identification by gait analysis and photogrammetry. J Forensic Sci 50(1):112–118
41. Murray MP (1967) Gait as a total pattern of movement. Am J Phys Med 46(1):290–333
42. Nilsson F et al (2008) Intelligent network video: understanding modern video surveillance systems. CRC Press
43. Nixon MS, Tan TN, Chellappa R (2005) Human identification based on gait. Springer, New York, Inc. Secaucus, NJ, USA
44. Niyogi SA, Adelson EH (1994) Analyzing and recognizing walking figures in XYT. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, pp 469–474
45. Rohr K (1994) Towards model-based recognition of human movements in image sequences. CVGIP Image Underst 59(1):94–115
46. Ross AA, Nandakumar K, Jain A (2006) Handbook of multibiometrics, vol 6. Springer Science & Business Media
47. Sarkar S, Phillips PJ, Liu Z, Vega IR, Grother P, Bowyer KW (2005) The humanID gait challenge problem: data sets, performance, and analysis. IEEE Trans Pattern Anal Mach Intel 27(2):162–177
48. Sivapalan S, Chen D, Denman S, Sridharan S, Fookes C (2011) Gait energy volumes and frontal gait recognition using depth images. In: International Joint Conference on, Biometrics (IJCB), 2011. IEEE, pp 1–6
49. Sun Y, Hare J, Nixon M (2016) Detecting acceleration for gait and crime scene analysis. http://eprints.soton.ac.uk/402396/
50. Tang J, Luo J, Tjahjadi T, Guo F (2017) Robust arbitrary-view gait recognition based on 3d partial similarity matching. IEEE Trans Image Process 26(1):7–22
51. Veres GV, Gordon L, Carter JN, Nixon MS (2004) What image information is important in silhouette-based gait recognition? In: Computer vision and pattern recognition, 2004, CVPR. IEEE computer society conference on, proceedings of the 2004. IEEE, vol 2, pp II–776
52. Wagg DK, Nixon MS (2004) On automated model-based extraction and analysis of gait. In: Proceedings of the sixth IEEE international conference on automatic face and gesture recognition, pp 11–16
53. Wang L, Ning H, Tan T, Hu W (2004) Fusion of static and dynamic body biometrics for gait recognition. IEEE Trans Circuits Syst Video Technol 14(2):149–158
54. Wei L, Tian Y, Wang Y, Huang T (2015) Swiss-system based cascade ranking for gait-based person re-identification. In: Twenty-ninth AAAI conference on artificial intelligence, pp 1882–1888
55. Wu Z, Huang Y, Wang L, Wang X, Tan T (2017) A comprehensive study on cross-view gait based human identification with deep CNNS. IEEE Trans Pattern Anal Mach Intel 39(2):209–226
56. Yam CY, Nixon M (2009) Gait recognition, model-based. In: Li S, Jain A (eds) Encyclopedia ofbiometrics. Springer, US, pp 633–639
57. Yam CY, Nixon MS, Carter JN (2004) Automated person recognition by walking and running via model-based approaches. Pattern Recognit 37(5):1057–1072
58. Yang SX, Larsen PK, Alkjær T, Simonsen EB, Lynnerup N (2013) Variability and similarity of gait as evaluated by joint angles: implications for forensic gait analysis. J Forensic Sci, pp 1556–4029

59. Zeng W, Wang C (2016) View-invariant gait recognition via deterministic learning. Neurocomputing 175:324–335
60. Zhao G, Liu G, Li H, Pietikainen M (2006) 3d gait recognition using multiple cameras. In: 7th international conference on automatic face and gesture recognition (FGR06). IEEE, pp 529–534

Chapter 2
Comparative Face Soft Biometrics for Human Identification

Nawaf Yousef Almudhahka, Mark S. Nixon and Jonathon S. Hare

Abstract The recent growth in CCTV systems and the challenges of automatically identifying humans under the adverse visual conditions of surveillance have increased the interest in soft biometrics, which are physical attributes that can be used to describe people semantically. Soft biometrics enable human identification based on verbal descriptions, and they can be captured in conditions where it is impossible to acquire traditional biometrics such as iris and fingerprint. The research on facial soft biometrics has tended to focus on identification using categorical attributes, whereas comparative attributes have shown a better accuracy. Nevertheless, the research in comparative facial soft biometrics has been limited to small constrained databases, while identification in surveillance systems involves unconstrained large databases. In this chapter, we explore human identification through comparative facial soft biometrics in large unconstrained databases using the Labelled Faces in the Wild (LFW) database. We propose a novel set of attributes and investigate their significance. Also, we analyse the reliability of comparative facial soft biometrics for realistic databases and explore identification and verification using comparative facial soft biometrics. The results of the performance analysis show that by comparing an unknown subject to a line up of ten subjects only; a correct match will be found in the top 2.08% retrieved subjects from a database of 4038 subjects.

Face Biometrics and Semantic Face Recognition

The crucial role of surveillance systems in crime prevention and public safety has motivated massive deployments of CCTV networks around the world [1–4].

N.Y. Almudhahka (✉) · M.S. Nixon · J.S. Hare
University of Southampton, Southampton, UK
e-mail: nya1g14@ecs.soton.ac.uk

M.S. Nixon
e-mail: msn@ecs.soton.ac.uk

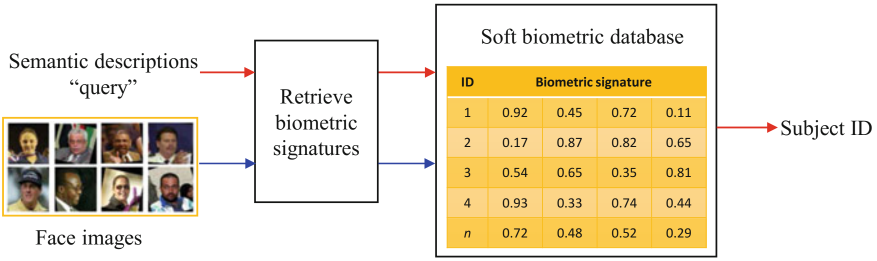
J.S. Hare
e-mail: jsh2@ecs.soton.ac.uk

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_2

For instance, the number of CCTV cameras deployed in cities and town centres in the UK was estimated between 1.85 [5] and 5.9 million [6]. This expansion in the usage of surveillance systems has increased the reliance on CCTV imagery for suspect identification, which is the first challenge faced by law enforcement agencies in criminal investigations [3]. Thus, the need for identifying suspects from imagery databases (i.e. mugshots or CCTV footage) has motivated research in human identification using semantic descriptions based on eyewitnesses' statements with a view to enabling searching a database of subjects through verbal descriptions [7–9]. These semantic descriptions are based on soft biometrics, which refer to physical and behavioural attributes that can be used to identify people.

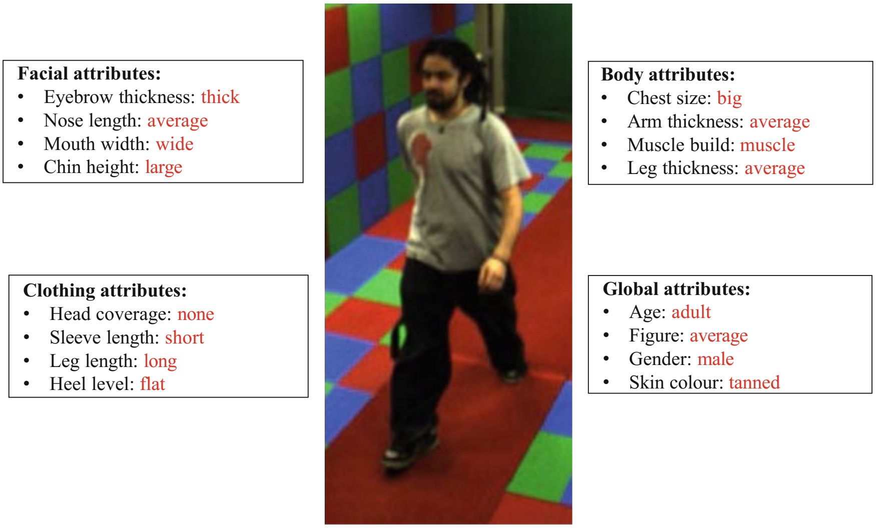
The main advantage of soft biometrics as compared with traditional hard biometrics (e.g. iris, DNA, and fingerprint) is that they can be acquired at a distance without individuals' involvement. In addition, soft biometrics enjoy sufficient robustness to the challenging visual conditions of surveillance such as occlusion of features, viewpoint variance, low resolution, and changes in illumination [7, 9, 10]. Therefore, soft biometrics can play a significant role in criminal investigations, where it is required to retrieve the identity of a suspect from an imagery database using a verbal description (i.e. eyewitness statement). Furthermore, soft biometrics bridge the semantic gap resulted from the difference between machines and humans in estimating attributes, which enables retrieval from a database solely by verbal descriptions as illustrated in Fig. 2.1.

Soft biometrics can be categorized according to their group and format. In terms of group, soft biometrics may fall under global, facial, body, or clothing attributes. While in terms of format, soft biometrics can be classified as: categorical, where individual attributes are assigned to specific classes (e.g. *square* versus *round* jaw); or comparative, where attributes of an individual are estimated relative to another individual (e.g. subject *A* has a *more rounded* jaw than subject *B*) [12]. Comparative soft biometrics are discussed further in more details in the next section. Figure 2.2 shows example categorical soft biometric attributes from the four different groups of soft biometrics: facial, body, clothing, and global. Figure 2.3 presents example soft biometric attributes in the comparative format. More highlights on the different soft biometric attribute formats are provided in Table 2.1.



ID	Biometric signature			
1	0.92	0.45	0.72	0.11
2	0.17	0.87	0.82	0.65
3	0.54	0.65	0.35	0.81
4	0.93	0.33	0.74	0.44
n	0.72	0.48	0.52	0.29

Fig. 2.1 Soft biometric identification using face attributes



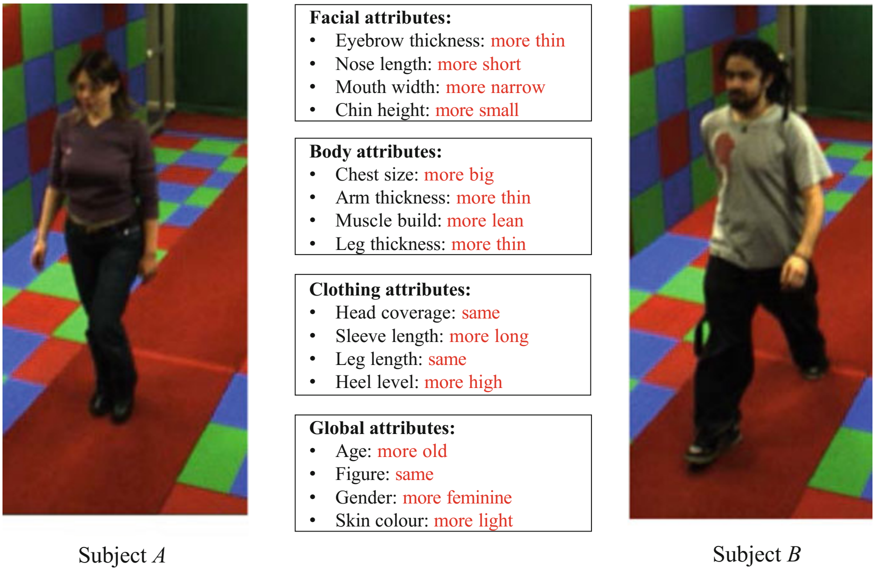
Facial attributes:
- Eyebrow thickness: thick
- Nose length: average
- Mouth width: wide
- Chin height: large

Body attributes:
- Chest size: big
- Arm thickness: average
- Muscle build: muscle
- Leg thickness: average

Clothing attributes:
- Head coverage: none
- Sleeve length: short
- Leg length: long
- Heel level: flat

Global attributes:
- Age: adult
- Figure: average
- Gender: male
- Skin colour: tanned

Fig. 2.2 Example categorical soft biometric attributes for a sample from the University of Southampton Biometric Tunnel (BioT) database [11]



Facial attributes:
- Eyebrow thickness: more thin
- Nose length: more short
- Mouth width: more narrow
- Chin height: more small

Body attributes:
- Chest size: more big
- Arm thickness: more thin
- Muscle build: more lean
- Leg thickness: more thin

Clothing attributes:
- Head coverage: same
- Sleeve length: more long
- Leg length: same
- Heel level: more high

Global attributes:
- Age: more old
- Figure: same
- Gender: more feminine
- Skin colour: more light

Subject *A*

Subject *B*

Fig. 2.3 Example comparative soft biometrics for a pair of subjects from the BioT database. The labels describe attributes of subject *A* with respect to subject *B*

Table 2.1 Semantic attribute formats

Format	Type	Nature of labels	Example
Categorical	Presence/absence	Binary (T/F)	*A* has **bushy** eyebrows
	Multiclass	Absolute	*A* has a **very large** mouth
Comparative	Similarity	=, ≠	*A* is the **same** age as *B*
	Dominance	<, >	*A* is **older** than *B*

Literature Review

Face Recognition in Psychology

Understanding face recognition from a psychological perspective is vital for studying facial identification using soft biometrics, as human identification accuracy is significantly affected by behavioural and physiological factors such as memory and encoding [13, 14]. In addition, human face recognition has substantial implications on automatic face recognition [15, 16], which has a wide range of real-life applications. The existing psychology literature on face recognition by humans is extensive and focuses particularly on understanding how the human visual and neural systems process the facial features and recognise faces. One of the key studies that has addressed face recognition in humans is that of Bruce and Young [17], who provisioned a theoretical model for understanding how humans recognize faces that is based on a set of distinguishable codes for face recognition, which are: label pictorial, structural, identity-specific semantic, visually derived semantic, and name. Bruce and Young's study suggested that regular face recognition involves the structural and identity-specific semantic codes (e.g. a familiar person's occupation and friends), while the other codes are used in lower levels for face recognition and perception. In [18], Hancock et al. outlined the factors that affect humans accuracy of recognizing unfamiliar faces. It has been found that changes in viewpoint and illumination significantly affect the ability of humans to recognize unfamiliar faces from images. Also, the spatial relationship between the facial components (i.e. configuration) has a high impact on recognition accuracy. Furthermore, the study emphasised the effect of face distinctiveness on recognition accuracy, and highlighted the role that can be played by the machines in aiding face recognition performance by humans.

A detailed overview on the human face recognition in a psychological and neural contexts was offered by O'Toole [19]. Her study provided some insights for addressing the problem of automatic face recognition. O'Toole stressed on humans ability to identify faces with the aid of semantic categories such as age, gender, and race, which increases recognition accuracy. Moreover, the study pointed the significance of the observer's as well as the observed person race on the recognition accuracy. Sinha et al. [15] outlined the key findings from previous experimental studies of face recognition in humans that can aid face recognition in computational systems. Their study highlighted the impact of spatial resolution on recognition performance, in addition to highlighting the holistic processing of facial features by the human visual system. Also, the study showed that pigmentation cues (i.e. texture or colour) have at least the same importance as the shape cues in face recognition. Tsao and Livingstone [20] highlighted the differences between computer vision algorithms and humans in face detection and recognition. The study emphasised on the independence of face detection and recognition stages in the human neural system, and outlined the differences in processing faces as compared with other objects in the neural system. Also, the study stressed the norm-based coding nature of face processing in humans, and the interpretation of faces in terms of their distinctiveness. Finally, the study highlighted the holistic processing of faces in the human neural system and stressed on the effect of the whole face on the processing of the individual facial components.

Several studies investigated the significance of facial features in recognition performance. Haig [21] explored the relative importance of facial features in face recognition by manipulating primary facial features of target samples and evaluating the recognition performance of the observers accordingly. The study found that eye-eyebrow, followed by mouth, and nose, have the highest importance for face recognition. Davies et al. [22] assessed the saliency of different facial features, where their experiments were based on manipulating faces using the Photofit Kit and monitoring the identification performance of different alternatives of samples. They have found that forehead and eyes have the highest saliency (i.e. their changes are most likely to be noticed), while chin and nose have the least saliency. Sadr et al. [23] investigated the role of eyes and eyebrows in face recognition and found that eyebrows have a significant role in face recognition that is at least similar to that of the eyes. Furthermore, their experiments reported that the absence of eyebrows results in a larger degradation in the face recognition performance as compared to the absence of eyes.

In summary, the findings of these studies highlight the role of semantic bindings in human face recognition, and show how categorization of facial or personal characteristics can affect the human face recognition performance. In addition, they outline the relative importance of the different facial parts, and show that eye-eyebrow region is the most important in face recognition. The implications of these studies will be noted on the proposed facial soft biometrics in this chapter.

Human Descriptions and Eyewitness Statements

There is a large volume of published studies that explored the reliability of verbal descriptions in eyewitness statements for suspects identification in criminal investigations. Kuehn [24] evaluated the effectiveness of verbal descriptions provided by victims of violent crimes to the police and examined the descriptions completeness using a random sample from the police records. The examined descriptions included the suspects' physical traits which are: age, sex, height, weight, build, skin colour, hair, and eye. The results of the study showed that more than 85% of the eyewitnesses were able to provide at least six attributes in their descriptions, where sex could be identified in 93% of the descriptions, while eye colour was the least to be recognised in the sample. The results also revealed that eyewitnesses cannot recall discrete traits of the suspects, but rather they have a general impression about the suspects.

In an analysis of the content of 139 eyewitnesses' descriptions, Sporer [25] found that 31% of the witnesses reported clothing attributes, 29.6% described facial features, 22% specified global features (i.e. age, and gender) in addition to movement descriptions, and the remaining descriptors used other features (e.g. smell, accent, and accessories). Sporer's analysis of the facial descriptions showed that most of the eyewitnesses described the upper half of the face, and more specifically, the hair of the suspect. Also, the study pointed that although the hair was the most mentioned in eyewitnesses descriptions, it is less reliable for locating the suspects as compared with the inner facial features, since hair style can be easily changed.

Koppen et al. [26] assessed the completeness and accuracy of eyewitnesses' descriptions using 2299 statements of 1313 eyewitnesses for 582 different robbers from official police records and investigated the factors that affect the accuracy and completeness of the statements. The findings emerged from their study revealed that the eyewitnesses tend to provide more information about general descriptions (e.g. age and gender) as compared to facial features. In addition, the study showed that the completeness of the descriptions did not necessarily imply their accuracy, thus although the information provided by the eyewitness was little, it tended to be accurate. In [27], Burton et al. explored subjects' ability to identify target people from surveillance videos. They found that familiarity with target faces has a substantial impact on the accuracy of identification. Thus, face recognition performance with familiar targets is much better than it is with unfamiliar ones. Furthermore, the study revealed that even with the poor quality of surveillance data, the face has a significantly higher impact on identification accuracy compared with the gait and body.

A significant analysis was presented by Meissner et al. [28], which outlined the psychological factors that affect eyewitness descriptions as: (1) encoding-based, which affects a person's perception such as illumination, distance, and stress; (2) person variables, which are age, gender, and race; and (3) the effect of inaccurate information from co-witnesses or investigators. Lee et al. [29] have conducted a detailed examination of the impact of a feature-based approach in suspect identification. Their experiments have shown that using a subjective feature-based approach for retrieving suspects from a database of mugshots is more efficient and accurate than presenting mugshots for an eyewitness in arbitrary order. Their experiments have also revealed that the feature-based approach of identifying suspects is effective for recognising faces in realistic conditions.

Overall, these studies reveal that the accuracy and completeness of eyewitnesses descriptions are determined by multiple factors such as the spatial and temporal settings of the incident, in addition to the eyewitness personal variables (e.g. age and gender). Furthermore, the findings of these studies stress on the tendency of eyewitnesses to describe general physical characteristics such as age and gender (i.e. global soft biometrics) in their statements, whereas facial features were described less likely. Also, the feature-based approach of identifying faces has been investigated and found to have a better impact on the efficiency and accuracy of suspects identification. Taken all together, these outcomes imply that global soft biometrics (e.g. age and gender) are essential for identification. In addition, the findings highlight the inadequacy of facial features for verbal descriptions as compared with other physical features, suggesting the introduction of more effective semantic facial attributes, which is one of the objectives of this chapter.

Facial Soft Biometrics

Due to its richness of features and details, the human face is considered as the most informative source of attributes for identification at a short distance as compared to other soft biometrics such as body and clothing [8, 9, 30]. Also, human face recognition demonstrated great robustness for challenging visual conditions such as low resolution and pose variability [31]. Therefore, a great deal of previous research into soft biometrics has focused on facial attributes either to improve the performance of traditional face recognition systems or to perform identification exclusively based on facial attributes [32].

The earliest exploration of semantic facial attributes emerged in [33], where face verification using the LFW database [34] was examined via attribute classifiers that were trained to recognize the presence or absence of a trait (i.e. categorical soft biometrics). The attributes covered 65 describable visual traits such as eyebrow shape, nose size, and eye width. The approach resulted in lowering the error rates by 23.92% compared to the state-of-the-art reported at that time on the LFW database. In [35], the authors studied the use of facial marks such as scars, moles, and freckles, as categorical soft biometrics to improve face recognition performance using the FERET database [36]. Their experiments demonstrated the improvement that can be achieved by augmenting facial marks (as soft biometrics) with traditional facial hard biometrics.

A key study in facial soft biometrics is that of Reid and Nixon [37], which is the first study to investigate human face identification using comparative soft biometrics. In this study, 27 comparative facial attributes were defined, and annotations were collected for subjects from the University of Southampton Gait Database (SGDB) [38]. The experiments showed that comparative facial soft biometrics outperform categorical facial soft biometrics. Thus, the rank-1 retrieval accuracy achieved using categorical attributes is 59.3%, compared to 74.5% in case of comparative attributes.

The first study that explored the interaction between automatically extracted soft biometrics and human generated soft biometrics is that of Klare et al. [39], which presented a method for using categorical facial attributes to perform identification in criminal investigations. Aiming to capture all persistent facial features, a set of 46 facial attributes were defined, and a model was trained to extract facial features and estimate them using SVM Regression automatically. Identification experiments were performed using the FERET [36] database with all the possible combination of probe-gallery (i.e. human vs. machine). Identification using an automatic probe and gallery resulted in the best recognition accuracy as compared with the other three identification scenarios in which human annotations are used (i.e. for probe, gallery, or both).

The study by Tome et al. [40] considered shape, orientation, and size of facial attributes as soft biometrics that can be utilized for forensic face recognition. The study proposed an approach to automatically convert a set of facial landmarks to a set of facial attributes (shape and size), which can be of continuous or discrete values (i.e. categorical). These features were used to generate statistics that aid forensic examiners in carrying morphological comparisons of facial images. Also, they were used to improve the performance of traditional face recognition system. Using ATVS [30] and MORPH [41] databases, the experiments revealed that the facial soft biometrics improve the accuracy of traditional face recognition systems. Recent work by Samangouei et al. [42] has investigated the use of categorical facial attributes for active authentication on mobile devices using the MOBIO [43] and AA01 [44] unconstrained mobile datasets. Their approach has highlighted the reliability of binary facial attributes for face verification on mobile devices and demonstrated the improvement that can be gained from fusing the scores of low-level features with the attribute-based approach. Table 2.2 summarizes the existing work that studied facial attributes for identification and verification.

In general, it can be seen from this literature review that the use of facial soft biometrics for human face identification has been studied using relatively constrained databases [8, 30, 39], whereas the real identification scenarios involve larger populations with significant demographic diversity, in addition to more challenging visual conditions of surveillance such as high variability in illumination, facial expressions, resolution, and pose. Except for Kumar et al. work [33], no study has so far addressed the use of facial soft biometrics for subject identification in large and unconstrained datasets like Labelled Faces in the Wild (LFW) [34]. In addition, although comparative soft biometrics have shown better identification accuracy as compared to categorical soft biometrics [8, 12], it is still not known whether comparative soft biometrics can scale for large and unconstrained datasets (e.g. LFW), as the studies in comparative soft biometric were performed using small and relatively

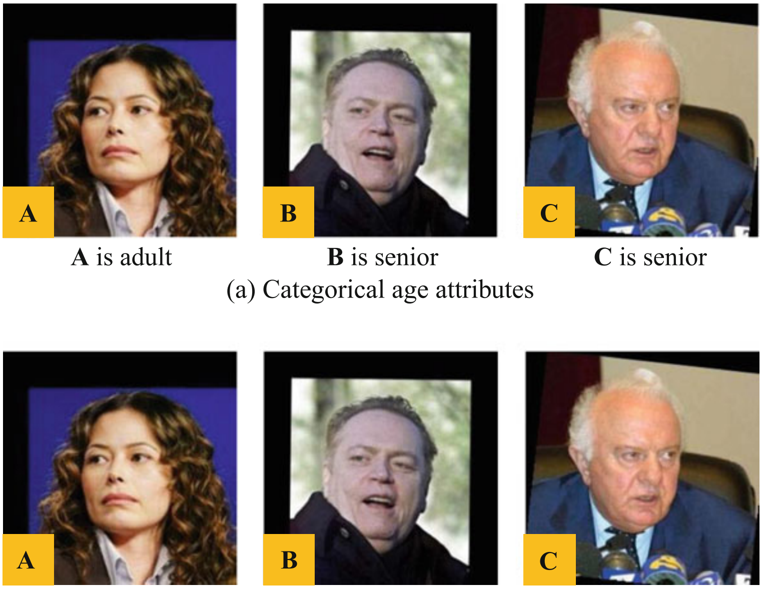
Table 2.2 Existing work on identification and verification using facial attributes

Publication	Dataset (no. of subjects)	Attributes	Labels
Kumar et al. [33]	LFW (5749)	65 categorical (binary)	Automatic
Reid and Nixon [37]	SGDB (100)	27 comparative	Human-based
Klare et al. [39]	FERET (1196)	46 categorical	Automatic and human-based
Tome et al. [40]	MORPH (130) and ATVS (50)	32 categorical	Automatic
Samangouei et al. [42]	MOBIO (152) and AA01 (50)	44 categorical (binary)	Automatic

constrained database [11]. Altogether, the findings from this literature review highlight the importance of exploring human identification using comparative facial soft biometrics in large unconstrained databases. Accordingly, this chapter aims to address the inadequacies of previous studies and to explores unconstrained human identification using comparative facial soft biometrics.

Comparative Facial Soft Biometrics

As mentioned in section "Face Biometrics and Semantic Face Recognition", the research in soft biometrics has largely been focused on categorical descriptions of facial attributes [7, 30, 39, 40, 45]. However, describing visual attributes in a relative (comparative) format has several advantages [46]. First, it makes richer semantics for humans (e.g. person *A* is *thinner* than person *B*). Second, it enables comparisons with a reference object (e.g. person *A* is *taller* than Bernie Ecclestone). Third, it improves interactive learning and makes searching based on an attribute more efficient (e.g. search for a *younger* person). Figure 2.4 illustrates the descriptive enrichment that can be gained by using comparative attributes. Besides these advantages of comparative attributes, the application of comparative soft biometrics in human identification



A is adult **B** is senior **C** is senior

(a) Categorical age attributes

A is younger than **B** **B** is younger than **C** **C** is older than **A**

(b) Comparative age attributes

Fig. 2.4 Expressing age of subjects from the LFW database [34, 47] using semantic labels

has demonstrated the superiority of relative attributes as compared to the categorical attributes [8, 10, 12].

The aim behind comparative soft biometrics is to create a biometric signature for each individual that embeds the individual's physical attributes, and consequently, allows each individual to be uniquely identified in a database as shown Fig. 2.1. This biometric signature is a vector that is composed of the relative strength of each soft biometric attribute. Relative rates are inferred from pairwise comparisons between the individual being evaluated and other individuals in the database. The generation of relative rates from pairwise comparisons can be achieved using a ranking algorithm that infers a rate for each item in a dataset from its pairwise comparisons. One popular ranking algorithm that has been used in prior work on comparative soft biometrics is the Elo rating system [8, 10, 12, 37], which is a well-known algorithm for rating chess players [48, 49]. Also, RankSVM [50], which is a formulation that learns a rating function from example pairwise comparisons to infer the relative rate of attributes, has been used in some studies on comparative soft biometrics [51, 52] to predict biometric signatures. The relative rating of soft biometric attributes using the Elo rating system is described later in this section.

Attributes and Comparative Labels

The human face has an abundance of features that can be used for identification. However, to consider a facial feature as an effective soft biometric attribute, it needs to be understandable, memorable, and describable. These aspects have governed the selection of facial features to define the soft biometric set used in this chapter, which covered the major facial components (i.e. eyes, eyebrows, mouth, and nose), with an emphasis on eyebrows according to their pivotal role in human face recognition [23]. Table 2.3 lists the proposed comparative soft biometric attributes that are analysed throughout this chapter. Each attribute is associated with a comparative label that is based on three-point bipolar scale, which ranges from -1 to 1, such that -1 is associated with the "*Less*" label, while 1 is associated with the "*More*" label. The normalised value of a comparative label is used to generate the relative rate of the attributes, which is computed using the Elo rating system as described later in section "Relative Rating of Attributes".

Dataset and Label Acquisition

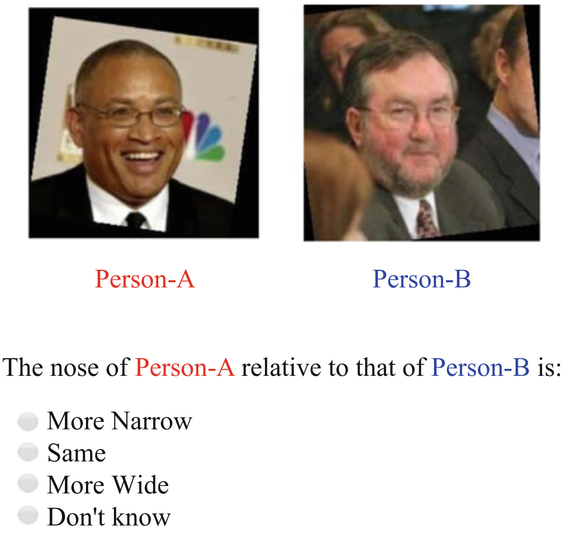
Labelled Faces in the Wild (LFW) [34] is a popular database that is used for unconstrained face recognition, and consists of more than 13000 facial images extracted from the web. The images of LFW have significant variations in pose, lighting, resolution, and facial expressions, which make them suitable to study unconstrained face recognition. The LFW database is composed of two subsets: *View 1*, which is

Table 2.3 Comparative facial soft biometrics defined to analyse unconstrained human identification and verification

No.	Attribute	Label
1	Chin height	[More Small, Same, More Large]
2	Eyebrow hair colour	[More Light, Same, More Dark]
3	Eyebrow length	[More Short, Same, More Long]
4	Eyebrow shape	[More Low, Same, More Raised]
5	Eyebrow thickness	[More Thin, Same, More Thick]
6	Eye-to-eyebrow distance	[More Small, Same, More Large]
7	Eye size	[More Small, Same, More Large]
8	Face length	[More Short, Same, More Long]
9	Face width	[More Narrow, Same, More Wide]
10	Facial hair	[Less Hair, Same, More Hair]
11	Forehead hair	[Less Hair, Same, More Hair]
12	Inter eyebrow distance	[More Small, Same, More Large]
13	Inter pupil distance	[More Small, Same, More Large]
14	Lips thickness	[More Thin, Same, More Thick]
15	Mouth width	[More Narrow, Same, More Wide]
16	Nose length	[More Short, Same, More Long]
17	Nose septum	[More Short, Same, More Long]
18	Nose-mouth distance	[More Short, Same, More Long]
19	Nose width	[More Narrow, Same, More Wide]
20	Spectacles	[Less Covered, Same, More Covered]
21	Age	[More Young, Same, More Old]
22	Figure	[More Thin, Same, More Thick]
23	Gender	[More Feminine, Same, More Masculine]
24	Skin colour	[More Light, Same, More Dark]

dedicated to training and model selection; and *View 2*, which is dedicated to performance analysis. The training subset of *View 1* consists of 9525 sample face images for 4038 subjects, some of these subjects have one sample in the database, while the others have two or more samples.

To explore unconstrained identification using comparative facial soft biometrics, a dataset that includes the 4038 subjects of the training subset of *View 1* from the LFW database was created by selecting one sample face image for each subject, and applying random selection whenever multiple samples exist for a subject. The selected images were all aligned using deep funnelling [47], which is an approach that incorporates unsupervised joint alignment with unsupervised feature learning to align face images, and reduces the effect of pose variability correspondingly. Also, all the images in the dataset were normalized to an inter-pupil distance of 50 pixels to ensure consistent comparisons between subjects.



Person-A

Person-B

The nose of Person-A relative to that of Person-B is:

○ More Narrow
○ Same
○ More Wide
○ Don't know

Fig. 2.5 Example crowdsourced comparison

The number of pairwise comparisons that result from a set of n items is $n(n-1)/2$, accordingly, the 4038 subjects in the LFW dataset result in 8.15 million pairwise comparisons, which is a massive number that is infeasible to be crowdsourced. Therefore, a graph that models pairwise relations between the 4038 subjects has been designed using a topology that ensures the involvement of each subject in at least four pairwise comparisons. The graph resulted in 10065 pairwise comparisons that were crowdsourced via the CrowdFlower platform,[1] and each of the 10065 crowdsourced comparisons consists of 24 questions targeting the comparative labelling of the attributes that are listed in Table 2.3. As explained earlier, each attribute is labelled based on a 3-point bipolar scale that represents the difference between the two subjects being compared. Figure 2.5 shows an example crowdsourced comparison. The crowdsourcing of the LFW dataset comparisons resulted in the collection of 241560 comparative labels for the 10065 comparisons as shown in Table 2.4. The labels collected through crowdsourcing were used to infer more comparisons. Thus, given two comparisons that involve subjects A and B with a common subject C, a new comparison between A and B can be inferred according to the rules outlined in Table 2.5. Relation inference results in increasing the coverage of the dataset, and enriches the analysis accordingly.

[1] http://www.crowdflower.com.

Table 2.4 Crowdsourcing job statistics

	Collected	Inferred	Total
Attribute comparisons	241560	132879504	133121064
Subject comparisons	10065	5536646	5546711
Average number of comparisons per subject	4.98	1371.1	N/A
Number of annotators (contributors)	9901	N/A	N/A

Table 2.5 Relation inference rules

(A, C)	(B, C)	inf(A, B)
=	=	=
>	<	>
<	>	<
>	=	>
<	=	<
>	>	N/A
<	<	N/A

Relative Rating of Attributes

Comparative soft biometrics aim to create a biometric signature for each individual that embeds the individual's physical attributes, and consequently, allows each individual to be uniquely identified. This biometric signature is a vector that is composed of the relative strength of each soft biometric attribute. The relative strength (or relative rate) is inferred from the pairwise comparisons between the individual begin evaluated and other individuals in the database. The generation of relative rates from pairwise comparisons can be achieved using a ranking algorithm that infers a rate for each item in a dataset from its pairwise comparisons. One popular ranking algorithm that has been used in prior work on comparative soft biometrics is the Elo rating system [8, 10, 12, 37], which is a well-known algorithm for rating chess players [48, 49]. Also, RankSVM [50], which is a formulation that learns a rating function from example pairwise comparisons to infer the relative rate of attributes, has been used in some studies on comparative soft biometrics [51, 52] to predict biometric signatures.

In this analysis, the Elo rating system is used to generate biometric signatures from comparisons, as its applicability and effectiveness for comparative soft biometrics have been already demonstrated [12, 37]. In addition, the Elo rating does not require training as is the case with RankSVM [50], which has also been proposed for rating comparative soft biometrics [51, 52]. The rating process in the Elo rating system starts by initializing the rates of all players in a tournament to an initial default value. Then, for a game between players A and B with the initial (default) rates R_A and R_B correspondingly, the expected scores, E_A and E_B are calculated as:

$$E_A = \left[1 + 10^{\frac{(R_B - R_A)}{400}}\right]^{-1} \tag{2.1}$$

$$E_B = \left[1 + 10^{\frac{(R_A - R_B)}{400}}\right]^{-1} \tag{2.2}$$

Subsequently, based on the game outcome (i.e. loss, win, or draw), the new rates, $\bar{R}_A$ and $\bar{R}_B$, for players A and B respectively, are:

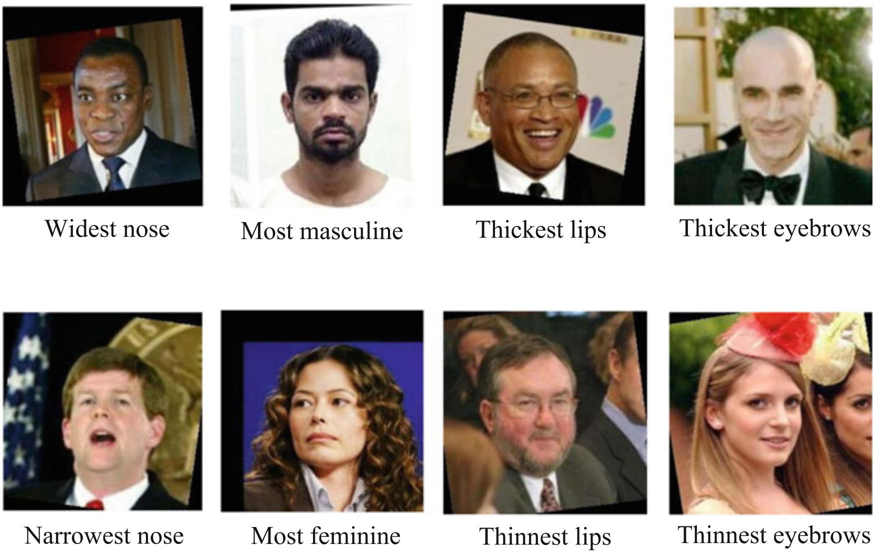
$$\bar{R}_A = R_A + K(S_A - E_A) \tag{2.3}$$

$$\bar{R}_B = R_B + K(S_B - E_B) \tag{2.4}$$

where S_A and S_B are scores that are set depending on the game outcome as: 0 for loss, 1 for win, and 0.5 for draw, while K is the score adjustment parameter that determines the sensitivity of rate update, and its value is selected through cross validation, as it has a significant impact on the outcomes from the rating process. The Elo rating system can be used to rate facial soft biometrics in a similar scheme to chess rating. Thus, by considering the subjects of a dataset as players in a tournament, and assuming that a comparison between two subjects, A and B, for a particular facial attribute, X, is a game between two players, correspondingly, this comparison can result in one of three possible outcomes for each of the two players as: "*Less X*", "*More X*", or "*Same X*", for example: "subject A has a *more thick* eyebrow than subject B". Accordingly, the rates of the two subjects that make the comparison are updated based on the comparison outcome using Eqs. 2.1–2.4. Figure 2.6 shows examples of the outcomes of the relative rating using the Elo system for selected attributes.

Attribute Analysis

The analysis in this section aim to explore three main aspects: (1) statistical distribution and the correlation between attributes; (2) attribute discriminative power; and (3) attribute semantic stability. The analysis in this section was conducted using the



Widest nose Most masculine Thickest lips Thickest eyebrows

Narrowest nose Most feminine Thinnest lips Thinnest eyebrows

Fig. 2.6 The subjects with the lowest and highest relative rates for selected attributes

relative rates of the attributes, which were generated from the crowdsourced comparative labels using the Elo rating system as explained in the previous section.

Statistical Overview

Fig. 2.7 shows the distribution of the attributes in the LFW dataset summarized using box plot. As can be noted from Fig. 2.7, the attributes significantly differ in their distribution and variation, which reflect the demographic diversity of the LFW dataset. Also, there is a strong presence of outliers with most of the attributes, which indicates the challenging nature of the dataset. Closer inspection of the distribution shows that *facial hair* and *spectacles*, which are more of a binary nature, have the greatest variation. On the contrary, *eyebrow hair colour* has the lowest variation. These findings might suggest the potential discriminative power of binary-like attributes, and the low power of *eyebrow hair colour*. This to be further investigated through the discriminative power analysis.

Investigating correlation between the attributes allows exploring their independence, in addition to the potential contribution of individual attributes in the distinguishing subjects. Figure 2.8 shows the correlation map for the attributes, where it can be seen that all correlations between the attributes are either insignificant or weak, with the exception of the negative correlation between *facial hair* and *skin colour*, which can be attributed to the low contrast between darker *skin colour* and

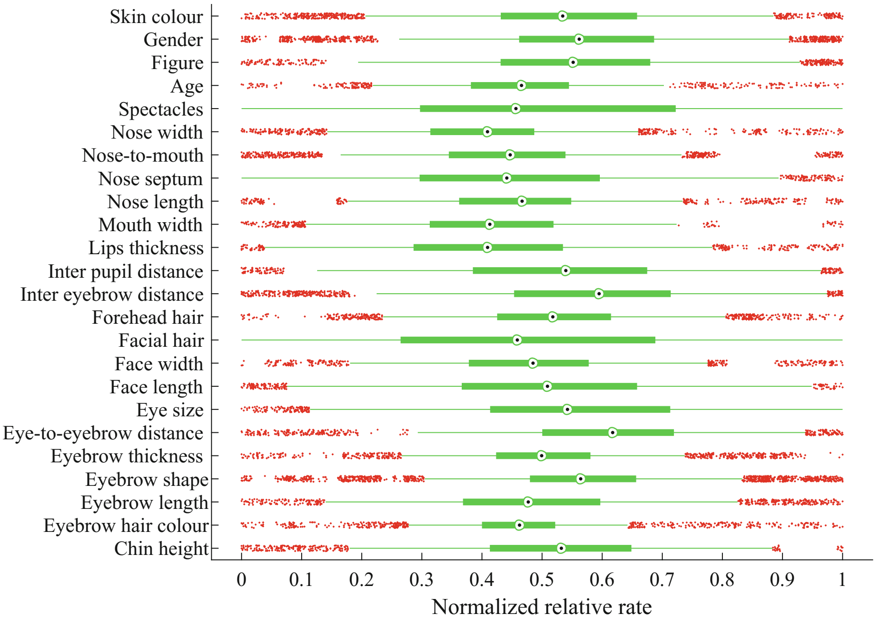


Fig. 2.7 Box plot for relative rates of the attributes

facial hair. In general, the findings from the correlation analysis confirm the independence of the attributes, and reveal the informative value embedded in each attribute.

Attribute Discriminative Power

Discriminative power analysis allows ranking attributes on their capabilities in distinguishing subjects, and contribution in identification accordingly. Discriminative power can be assessed using entropy [39], which is an information theoretic measure that represents the average amount of information contained in a random variable X, and it is calculated as follows:

$$H(X) = -\sum_{x \in X} p(x) log_2 \left[p(x) \right] \tag{2.5}$$

where X is a discrete random variable, and $p(x)$ is the probability distribution function of X. In the context of soft biometrics, it is assumed that the relative rates of attributes are random variables, and thus, entropy can be used to measure the information contained in each attribute, providing us with an indicator of the impact of each attribute in distinguishing between subjects.

Figure 2.9 shows the discriminative power of the attributes in terms of normalized entropy. It can be seen from Fig. 2.9 that *spectacles*, *facial hair*, and *gender*, which

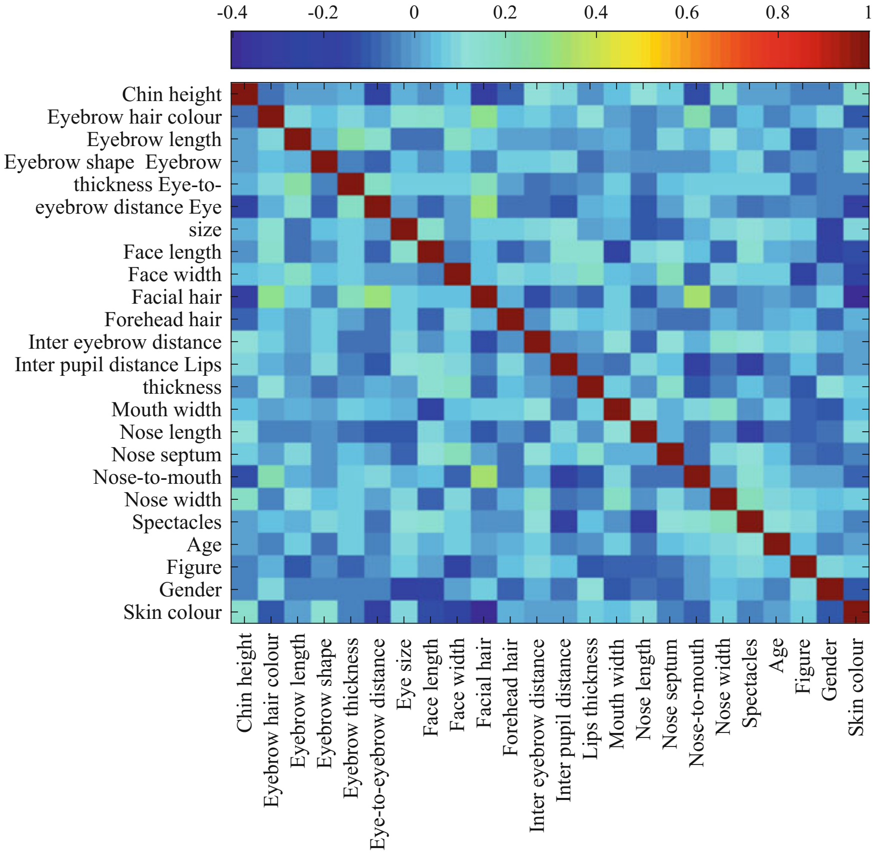


Fig. 2.8 Correlations between the attributes

are binary-like attributes, have relatively high discriminative power. Also, the results show the relatively high discriminative power of *eyebrow shape* and *inter eyebrow distance*, which highlight the role of eyebrows in human face recognition. On the other hand, the analysis shows that *age* has the lowest discriminative power, which can be attributed to the inaccuracy of human estimations for age from face images [53]. Furthermore, *eyebrow hair colour* and *forehead hair* demonstrate low discriminative power, which can be due to the difficulties of estimating these attributes.

Attribute Semantic Stability

Semantic stability can be defined as the consistency of an attribute rate among different annotators, which is substantial in assessing the attribute effectiveness and

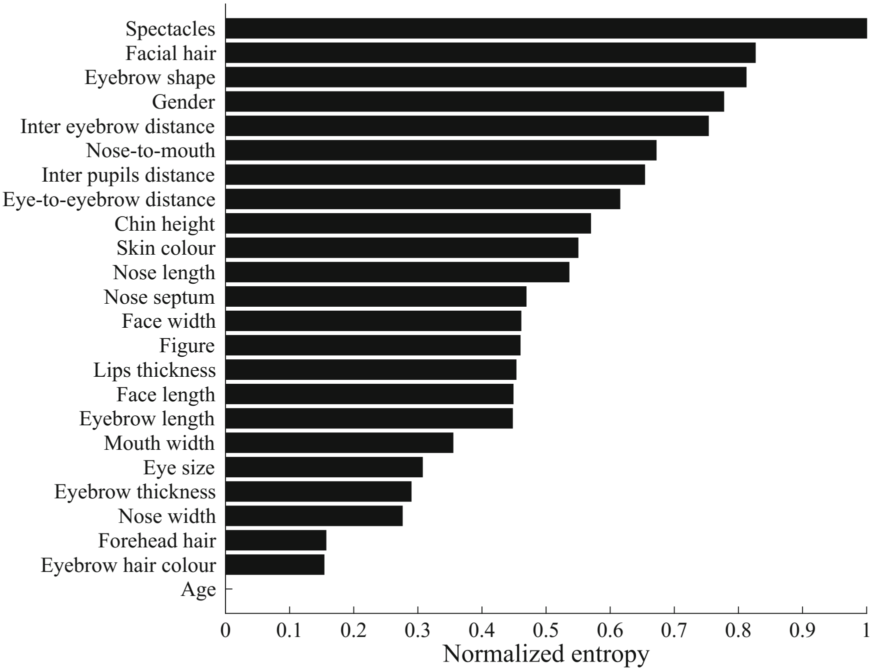


Fig. 2.9 Discriminative power of the attributes

robustness. The semantic stability of the attributes is evaluated by creating two different galleries, each of which consists of the biometric signatures of all the subjects in the dataset, where each biometric signature is composed of the relative rates of the 24 soft biometric attributes (listed in Table 2.3). The relative rates in each gallery are inferred using the Elo rating system based on two mutually exclusive subsets of comparative labels, which represent two different groups of annotators. Then, the semantic stability is measured for each attribute among the two galleries as the Pearson's correlation between the subjects' rates in both galleries. The results of the semantic stability analysis are shown in Fig. 2.10.

The semantic stability analysis demonstrated that all the attributes are statistically significant ($p < 0.05$), regardless of the strength of the correlation (stability) between the two galleries. Moreover, an interesting outcome from the result of the semantic stability analysis is the agreement of its ranking for the attributes with the ranking resulted from the discriminative power analysis in Fig. 2.9. Thus, the binary-like attributes (i.e. *spectacles*, *facial hair*, and *gender*) beside *eyebrow shape* have the highest semantic stability, while *age*, *eyebrow hair colour*, and *forehead hair* have the lowest discriminative power. This correspondence in the findings of semantic stability and discriminative power analysis reveals the robustness of entropy for assessing discriminative power of soft biometric attributes.

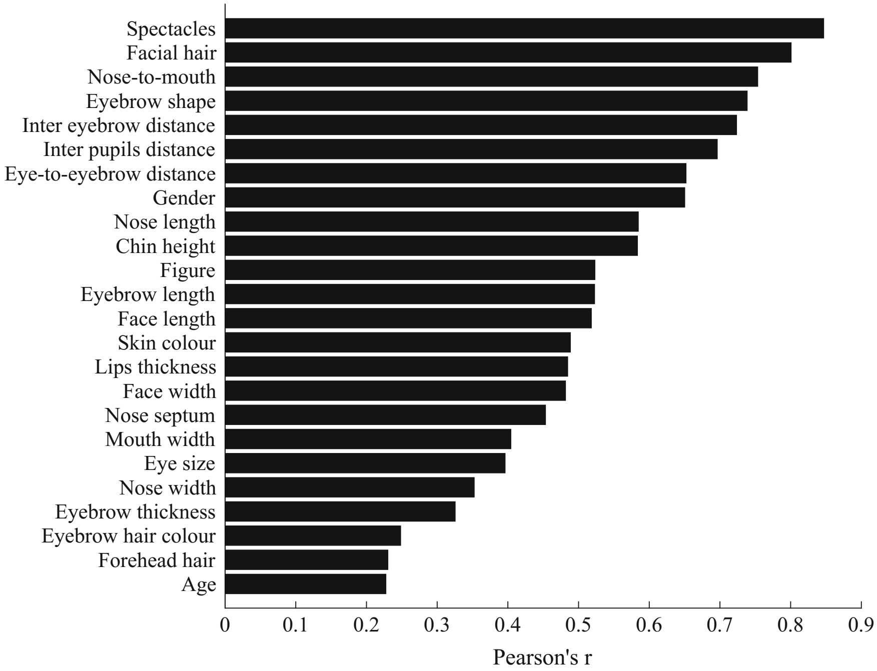


Fig. 2.10 Semantic stability of the attributes

Analysing Facial Comparisons

Identification Using Facial Comparisons

This experiment simulates a realistic scenario in which a semantic database DB_s is searched to identify an unknown subject using a verbal description for the subject's face as illustrated in Fig. 2.1. The identification performance evaluation follows a 6-fold cross validation, where the 4038 subjects of the LFW dataset are randomly divided into six equal subsets, and each subset is used for testing while the remaining five folds are used for training. For each subject in the test set, a probe biometric signature PR_s is generated from the relative rates of the 24 attributes (listed in Table 2.3). The relative rates are computed using the Elo rating system and based on comparisons between the probe and c other randomly selected subjects from the training folds. The remaining comparisons, after excluding those used in generating PR_s, are used to generate a biometric signature for each subject, which makes up the database DB_s to be searched. Then, the distance, d_p, between the probe and each subject in DB_s is calculated using the Pearson correlation coefficient as follows:

$$d_p = 1 - \frac{\sum_{i=1}^{t}(PR_s(i) - \overline{PR}_s)(S_c(i) - \overline{S}_c)}{\sqrt{\sum_{i=1}^{t}(PR_s(i) - \overline{PR}_s)^2}\sqrt{\sum_{i=1}^{t}(S_c(i) - \overline{S}_c)^2}} \tag{2.6}$$

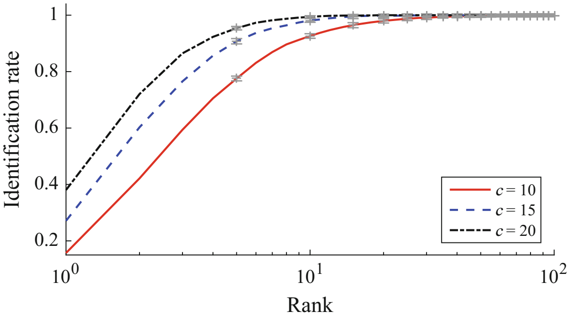
where PR_s is the biometric signature of the probe, $S_c \in DB_s$ is the biometric signature of the counterpart subject in the database DB_s that is compared to the probe, and t is the number of features composing the semantic face signature. The rank of the correct match to the probe is used to report the identification performance via a Cumulative Match Characteristic (CMC) curve. This cross validation runs over the six folds, and repeated till the harmonic mean of identification rates among all the ranks converges. Figure 2.11 shows the CMC curve resulted from this experiment and it can be seen that using ten subject comparisons to generate the probe biometric signature, which is the ideal size of identity parade [10], an identification rate of 92.62% can be achieved. Rank-10 identification rate increases to 98.14% and 99.41% as the number of subject comparisons increase to 15 and 20 respectively.

The relationship between the number of comparisons used in generating biometric signatures and identification performance can be seen in Fig. 2.12 and it shows that the identification performance in terms of mean rank of retrieved match converges starting from eight comparisons. Furthermore, the impact of using facial comparisons on identification performance can also be seen from another interesting perspective, which is the compression of search range in a database. Thus, narrowing down search range becomes vital for efficiency of identification in large database. In addition, when verbal descriptions are not sufficiently accurate, search compression can lead to filtering out a long list of suspects, making subject retrieval more efficient.

Figure 2.13 demonstrates the compression in the search range that can be achieved in the LFW dataset using the comparative facial soft biometrics. It shows that using two comparisons only, the search range can be narrowed down to 10.55% of the total dataset with probability $p = 0.99$ that a correct match with the subject will be found.

To the best of our knowledge, the only study that has investigated human identification using facial soft biometrics for both probe and gallery in a relatively large database is that of Klare et al. [39], which achieved a rank-1 accuracy of 22.5% using

Fig. 2.11 Identification performance using $c = \{10, 15, 20\}$ subject comparisons



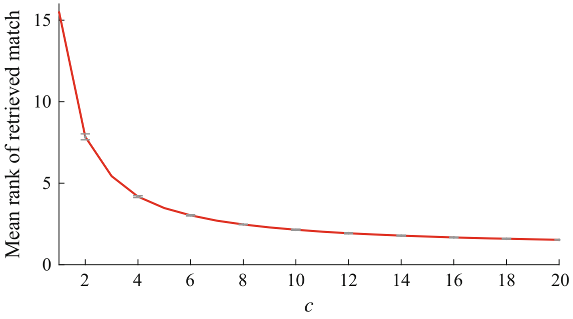


Fig. 2.12 Effect of number of subject comparisons, *c*, on identification performance

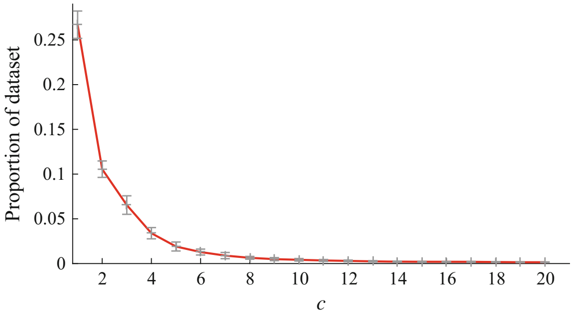


Fig. 2.13 Compression achieved in search range with probability of finding a correct match $p = 0.99$ at different number of comparisons, *c*

46 attributes (27 categorical and 19 binary) with 1196 subjects from the relatively constrained FERET database. Our approach can reach a rank-1 accuracy of 38.04% with 20 subject comparisons and 24 comparative attributes only, using the larger and more challenging LFW dataset. This performance advantage is due to the use of comparative attributes [10]. Furthermore, the results of the identification experiment demonstrate the power of comparative soft biometrics for unconstrained face identification, and reveal their scalability for relatively large databases.

Verification Using Facial Comparisons

Analysing the verification accuracy of comparative facial soft biometrics is necessary to evaluate the extent of agreement among semantic descriptions collected from different annotators (i.e. eyewitnesses) for the same subject. Each biometric signature for a subject can be considered as a unique sample, provided that it was generated using a unique group of comparisons. Therefore, in this experiment, two samples (biometric signatures) were generated for each of the 4038 subjects of the LFW dataset using randomly selected and mutually exclusive comparisons. The number of comparisons that were selected to generate biometric signatures was set to 10, as it is the average ideal size for an identity parade [10]. The verification was assessed using the two galleries resulted from the generated biometric signatures.

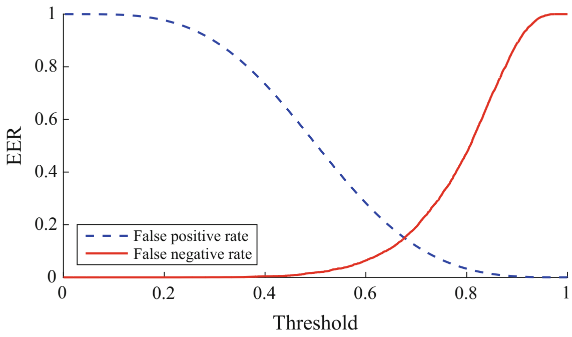
As Fig. 2.14 shows, the comparative facial soft biometrics achieved an equal error rate (EER) of 15.4% using ten subject comparisons only. Also, the attributes achieved an area under the curve (AUC) of 92.32% as can be seen in Fig. 2.15a. Furthermore, it can be seen from Fig. 2.15b that as the number of subjects comparisons used in generating biometric signature increases, the verification performances improves.

Using the BioT database, Tome et al. [30] achieved an EER of 13.54% by utilizing an automatic face detection and recognition system, with the fusion of 23 categorical soft biometrics (13 bodily, 3 global, and 7 head). While the proposed comparative attributes (without face hard biometrics) resulted in a slightly higher EER of 15.4% using ten subject comparisons, which decreases to 11.32% outperforming the approach in [30] with five more subjects comparisons only. Overall, the results of this experiment demonstrate the verification capability of the proposed comparative soft biometrics and show that the attributes can outperform the performance of automatic face recognition with or without fusing categorical soft biometrics.

Approach Limitations

Although the effectiveness and reliability of the proposed approach have been demonstrated in this chapter, some issues need to be considered for the practical application of the approach. First, the subject comparisons have taken place while the images of subject pairs were presented for the annotators, whereas real identification scenarios involve recalling suspects' facial attributes from eyewitnesses' memory. Second, as explained in section "Relative Rating of Attributes", the relative rating of comparative soft biometrics using the Elo rating system involves a cross validation to tune the score adjustment parameter, and this can be costly with large databases. Third, despite its effectiveness for studying comparative soft biometrics, crowdsourcing of comparisons is not scalable for very large databases due to the time and cost associated with it. Accordingly, a framework for the automatic estimation of

Fig. 2.14 Error curves resulted from verification using facial comparisons



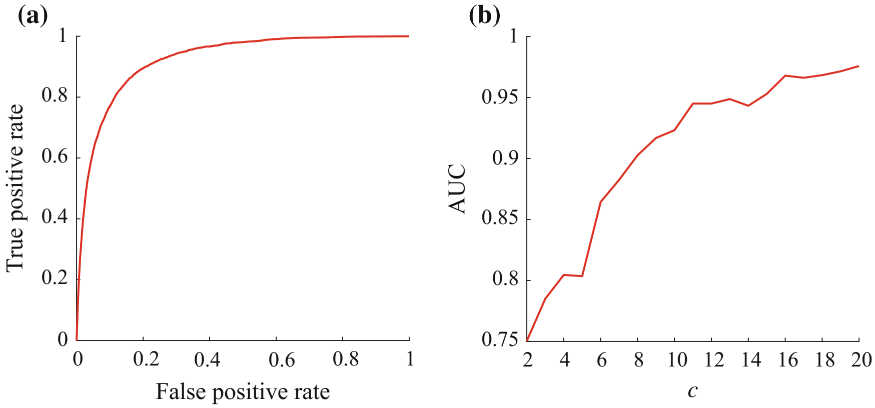


Fig. 2.15 **a** ROC curve for comparative facial attributes using $c = 10$ subject comparisons. **b** Area under the curve (AUC) versus number of subject comparisons, c

comparative facial soft biometrics is required to reduce the dependency on human annotators. Finally, most of the subjects of the LFW database are public figures that are familiar to the annotators, while the real life scenarios involve the description of unfamiliar faces (e.g. eyewitness descriptions of a suspects), which might be more difficult to recall and describe.

Summary

This chapter aims to analyse unconstrained human identification using comparative facial soft biometrics. We have presented a literature review on identification using facial attributes and highlighted the need to bridge the knowledge gap in comparative facial soft biometrics. Also, we have proposed a novel set of comparative facial soft biometrics and demonstrated their significance. The performance analysis conducted in this chapter using the well-known LFW database have revealed the reliability of comparative facial soft biometrics for unconstrained identification and verification, in addition to their scalability for large databases. Thus, by comparing an unknown subject to a line up of ten subjects only, a correct match will be found in the top 84 retrieved subjects from a database of 4038 subjects. Furthermore, the attributes have revealed a notable verification accuracy achieving an EER of 15.4% with biometric signatures generated from ten subject comparisons only. In conclusion, the findings of this chapter extend our knowledge of capabilities of comparative soft biometrics and can serve as a base for future investigations of other comparative soft biometrics such as body and clothing.

In terms of directions for future work, it would be interesting to explore the automatic retrieval of comparative facial soft biometrics from images. Also, as criminal investigations involve the collection of memory-based descriptions from eyewitnesses, future work could determine the accuracy of humans in recalling facial attributes of unknown subjects from memory relative to presented subjects. and exploring the effectiveness of these soft biometric comparisons for identity retrieval.

References

1. Davis JP, Valentine T (2015) Human verification of identity from photographic images. In: Forensic facial identification: theory and practice of identification from eyewitnesses, composites and CCTV, pp 209–238
2. Crawford TAM, Evans K (2016) Crime prevention and community safety. Oxford University Press
3. Valentine T, Davis JP (2015) Forensic facial identification. In: Forensic facial identification: theory and practice of identification from eyewitnesses, composites and CCTV, pp 323–347
4. Kemp R, White D (2016) Face identification 3. In: An introduction to applied cognitive psychology, pp 39
5. Gerrard G, Thompson R (2011) Two million cameras in the UK. CCTV Image 42(10):e2
6. Barrett D (2013) One surveillance camera for every 11 people in Britain, says CCTV survey. The Telegraph 10
7. Samangooei S, Guo B, Nixon MS (2008) The use of semantic human description as a soft biometric. In: 2008 2nd IEEE international conference on biometrics: theory, applications and systems, BTAS 2008. IEEE, pp 1–7
8. Reid D, Samangooei S, Chen C, Nixon M, Ross A (2013) Soft biometrics for surveillance: an overview. Mach Learn Theory Appl 327–352
9. Nixon MS, Correia PL, Nasrollahi K, Moeslund TB, Hadid A, Tistarelli M (2015) On soft biometrics. Pattern Recognit Lett 68:218–230
10. Reid DA, Nixon MS, Stevenage SV (2014) Soft biometrics; human identification using comparative descriptions. IEEE Trans Pattern Anal Mach Intell 36(6):1216–1228
11. Seely RD, Samangooei S, Lee M, Carter JN, Nixon MS (2008) The University of Southampton multi-biometric tunnel and introducing a novel 3d gait dataset. In 2008 2nd IEEE international conference on biometrics: theory, applications and systems, BTAS 2008. IEEE, pp 1–6
12. Reid DA, Nixon MS (2011) Using comparative human descriptions for soft biometrics. In: 2011 international joint conference on biometrics (IJCB). IEEE, pp 1–6
13. Ryder H, Smith HMJ, Flowe HD (2015) Estimator variables and memory for faces. In: Forensic facial identification: theory and practice of identification from eyewitnesses, composites and CCTV, pp 159–183
14. Robbins R, McKone E (2007) No face-like processing for objects-of-expertise in three behavioural tasks. Cognition 103(1):34–79
15. Sinha P, Balas B, Ostrovsky Y, Russell R (2006) Face recognition by humans: nineteen results all computer vision researchers should know about. Proc IEEE 94(11):1948–1962
16. Jonathon P (2014) Phillips and Alice J O'Toole. Comparison of human and computer performance across face recognition experiments. Image Vis Comput 32(1):74–85
17. Bruce V, Young A (1986) Understanding face recognition. Br J Psychol 77(3):305–327
18. Hancock PJB, Bruce V, Burton AM (2000) Recognition of unfamiliar faces. Trends Cogn Sci 4(9):330–337
19. O'Toole AJ (2005) Psychological and neural perspectives on human face recognition. In: Handbook of face recognition. Springer, pp 349–369
20. Tsao DY, Livingstone MS (2008) Mechanisms of face perception. Annu Rev Neurosci 31:411–437
21. Nigel D (1986) Haig. Exploring recognition with interchanged facial features. Perception 15(3):235–247

22. Davies G, Ellis H, Shepherd J (1977) Cue saliency in faces as assessed by the 'Photofit' technique. Perception 6(3):263–269
23. Sadr J, Jarudi I, Sinha P (2003) The role of eyebrows in face recognition. Perception 32(3):285–293
24. Lowell L (1974) Kuehn. Looking down a gun barrel: Person perception and violent crime. Percept Mot Skills 39(3):1159–1164
25. Sporer SL (1992) An archival analysis of person descriptions. In: Biennial meeting of the American Psychology-Law Society in San Diego, California
26. PJ van Koppen, Lochun SK (1997) Portraying perpetrators: The validity of offender descriptions by witnesses. Law Hum Behav 21(6):661
27. Burton MA, Wilson S, Cowan M, Bruce V (1999) Face recognition in poor-quality video: Evidence from security surveillance. Psychol Sci 10(3):243–248
28. Meissner CA, Sporer SL, Schooler JW (2013) Person descriptions as eyewitness evidence. In: Handbook of eyewitness psychology: memory for people, pp 1–34
29. Lee E, Whalen T, Sakalauskas J, Baigent G, Bisesar C, McCarthy A, Reid G, Wotton C (2004) Suspect identification by facial features. Ergonomics 47(7):719–747
30. Tome P, Fierrez J, Vera-Rodriguez R, Nixon MS (2014) Soft biometrics and their application in person recognition at a distance. IEEE Trans Inf Forensics Secur 9(3):464–475
31. O'Toole A, Jonathon Phillips P (2015) Evaluating automatic face recognition systems with human benchmarks. In: Forensic facial identification: theory and practice of identification from eyewitnesses, composites and CCTV, p 263
32. Arigbabu OA, Ahmad SM, Adnan WA, Yussof S (2015) Sharifah Mumtazah Ahmad, Wan Azizun Adnan, and Salman Yussof. (2015) Recent advances in facial soft biometrics. Vis Comput Int J Comput Graph 31(5):513–525
33. Kumar N, Berg AC, Belhumeur PN, Nayar SK (2009) Attribute and simile classifiers for face verification. In: 2009 IEEE 12th International Conference on Computer Vision, IEEE, pp 365–372
34. Huang GB, Ramesh M, Berg T, Learned-Miller E (2007) Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical report, Technical report 07–49, University of Massachusetts, Amherst
35. Park U, Jain AK (2010) Face matching and retrieval using soft biometrics. IEEE Trans Inform Forensics Secur 5(3):406–415
36. Phillips JP, Moon H, Rizvi SA, Rauss PJ (2000) The feret evaluation methodology for face-recognition algorithms. IEEE Trans Pattern Anal Mach Intell 22(10):1090–1104
37. Reid DA, Nixon MS (2013) Human identification using facial comparative descriptions. In: 2013 International Conference on Biometrics (ICB). IEEE, pp 1–7
38. Shutler JD, Grant MG, Nixon MS, Carter JN (2004) On a large sequence-based human gait database. In: Applications and science in soft computing. Springer, pp 339–346
39. Klare BF, Klum S, Klontz JC, Taborsky E, Akgul T, Jain AK (2014) Suspect identification based on descriptive facial attributes. In: 2014 IEEE International Joint Conference on Biometrics (IJCB). IEEE, pp 1–8
40. Tome P, Vera-Rodriguez R, Fierrez J, Ortega-Garcia J (2015) Facial soft biometric features for forensic face recognition. Forensic Sci Int 257:271–284
41. Ricanek K, Tesafaye T (2006) Morph: a longitudinal image database of normal adult age-progression. In: 2006 7th international conference on automatic face and gesture recognition, FGR 2006. IEEE, pp 341–345
42. Samangouei P, Patel VM, Chellappa R (2017) Facial attributes for active authentication on mobile devices. Image Vis Comput 58:181–192
43. McCool C, Marcel S, Hadid A, Pietikäinen M, Matejka P, Cernockỳ J, Poh N, Kittler J, Larcher A, Levy C et al (2012) Bi-modal person recognition on a mobile phone: using mobile phone data. In: 2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). IEEE, pp 635–640
44. Fathy ME Patel VM, Chellappa R (2015) Face-based active authentication on mobile devices. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp 1687–1691

45. Jain AK, Park U (2009) Facial marks: soft biometric for face recognition. In: 2009 16th IEEE International Conference on Image Processing (ICIP). IEEE, pp 37–40
46. Parikh D, Grauman K (2011) Relative attributes. In: 2011 IEEE International conference on computer vision (ICCV). IEEE, pp 503–510
47. Huang G, Mattar M, Lee H, Learned-Miller EG (2012) Learning to align from scratch. In: Advances in neural information processing systems, pp 764–772
48. Glickman ME (1995) A comprehensive guide to chess ratings. Am Chess J 3:59–102
49. Elo AE (1978) The rating of chessplayers, past and present. Arco Publishing
50. Joachims T (2002) Optimizing search engines using clickthrough data. In: Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 133–142
51. Jaha ES, Nixon MS (2014) Soft biometrics for subject identification using clothing attributes. In: 2014 IEEE international joint conference on biometrics (IJCB). IEEE, pp 1–6
52. Martinho-Corbishley D, Nixon MS, Carter JN (2016) Soft biometric recognition from comparative crowdsourced annotations. IET Biometrics 1–16
53. Han H, Jain AK (2014) Age, gender and race estimation from unconstrained face images. Department of Computer Science Engineering, Michigan State University, East Lansing, MI, USA, MSU Technical Report (MSU-CSE-14-5)

Chapter 3
Video-Based Human Respiratory Wavelet Extraction and Identity Recognition

Xue Yang and Thirimachos Bourlai

Abstract In this paper, we study the problem of human identity recognition using off-angle human faces. Our proposed system is composed of (i) a physiology-based human clustering module and (ii) an identification module based on facial features (nose, mouth, etc.) fetched from face videos. In our proposed methodology we, first, passively extract an important vital sign (breath). Next we cluster human subjects into nostril motion versus nostril non-motion groups, and, then, localize a set of facial features, before we apply feature extraction and matching. Our proposed human identity recognition system is very efficient. It is working well when dealing with breath signals and a combination of different facial components acquired under challenging luminous conditions. This is achieved by using our proposed Motion Classification approach and Feature Clustering technique based on the breathing waveforms we produce. The contributions of this work are three-fold. First, we generated a set of different datasets where we tested our proposed approach. Specifically, we considered six different types of facial components and their combination, to generate face-based video datasets, which present two diverse data collection conditions, i.e., videos acquired in head full frontal pose (baseline) and head looking up pose. Second, we propose an alternative way of passively measuring human breath from face videos. We demonstrate a comparatively identical breath waveform estimation when compared against the breath waveforms produced by an ADInstruments device (baseline) (Adinstruments, http://www.adinstruments.com/ [1]). Third, we demonstrate good human recognition performance based on partial facial features when using the proposed pre-processing Motion Classification and Feature Clustering techniques. Our approach achieves increased identification rates across all datasets used, and it yields a significantly high identification rate, ranging from 96 to 100% when using a single or a combination of facial features. The approach yields an average of 7% rank-1 rate increase, when compared to the baseline scenario. To the best of our knowledge, this is the first time that a biometric recognition system positively exploits human

X. Yang (✉) · T. Bourlai
West Virginia University, PO Box 6201, Morgantown, USA
e-mail: yangxue0629@gmail.com

T. Bourlai
e-mail: thirimachos.bourlai@mail.wvu.edu

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_3

breath waveforms, which when fused with partial facial features, it increases a benchmark face-based recognition performance established using academic face matching algorithms.

Introduction

Identity security is becoming increasingly important for our society. Nowadays, there are examples of real-life cases indicating that the traditional authentication techniques based on passwords, identity documents or physical tokens fail to provide enough safety. Therefore, in order to authenticate personal identity, it would be more convenient and secure for users to employ a set of biometric features derived directly from their physical or behavioral characteristics (also known as traits or identifiers) [2]. There already exists several biometric authorization systems [3] or simple biometric features (such as fingerprint [4], face [5], hand geometry [6], iris [7], signature [8], etc.) in use for diverse applications [9]. Compared to traditional authorization schemes, biometric-based schemes use biometric traits which are more reliable, as they cannot be forgotten or lost. It is also difficult for others to share or duplicate the inherent characteristics of human beings [2]. We also see that the security focused industry promotes further research on the development of novel biometric systems that can help supplement or replace the traditional ways of human authentication.. For example, lately we have seen a number of studies demonstrating the capability of using the electrocardiogram (ECG) signal for human identity recognition [10–13].

Inspired by the analysis of ECG-based identification, we propose a novel video-based biometric approach that is composed of (i) a physiology-based human clustering module (passive respiratory monitoring using a visible camera) and (ii) an identification module based on facial features (nose, mouth, etc.) fetched from face videos. A typical application of such an approach is login access on portable devices (laptops, cell phones etc.) using face recognition (FR). While several techniques are used for respiratory monitoring [14], such as spirometers [15], nasal thermo-couples [16], transthoracic inductance, impedance plethysmography [17], strain gauge, etc., each of these techniques requires a dedicated device to be attached to some part of the human body. Our proposed method is completely passive. It uses the clusters of similar respiratory wavelets extracted directly from videos of human faces to improve face-based recognition performance.

The face recognition component-based approach used can be beneficial on various security applications [18]. The main benefit of using components is the fact that, depending on pose and quality, single or multiple facial features can be used for human recognition [19]. In [18] face recognition studies were conducted by combining a set of facial components into a single feature vector and classifying using linear Support Vector Machines (SVM). An improved technique was proposed by combining 3D morphable models with component-based recognition. The technique computes the 3D face models from input face images, which are used for the training step of the proposed component-based FR system [20]. Also some previous studies

[19, 21] discuss a set of approaches that propose the selection of discriminatory facial components to categorize subjects. There are also other spectral independent FR component-based approaches proposed in the literature, where face matching performance, under certain conditions (e.g., on images with high image and face-based quality) was reported to be high [22]. For example, FR in the mid-wave infrared (MWIR) band [23]. In some of these band specific FR approaches, matching is performed utilizing fiducial points, when using either the whole region or sub-regions of the human face.

In our study, we mainly focus on human face identification using the nose region in combination with other facial components (e.g., eyes, mouth). The pre-processing stage for our component-based FR system is composed of a physiology-based classification module, where the features used for classification are generated by respiratory wavelets. In this module the subjects are automatically grouped into either a nostril motion or a nostril non-motion group before matching is performed. The purpose of the classification module is to improve the identification performance of such a hybrid biometric system. This will be discussed in more detail in the following sections.

Goals and Contributions:

The goals of this work include (1) the design of a non-contact measurement of the human breathing signal based on face-based recorded videos, and (2) the development of a pre-processing methodology that can accurately group each of the original biometric gallery and probe datasets into two smaller subsets (nostril-motion specific cohorts), prior to applying single or multi-facial component-based matching, so that our established benchmark face identification performance is improved.

In this regard, we *first* propose a respiratory extraction schema. The objective is, for each frame of our available subject-specific face video, to detect and localize the nose and each nostril, and then, measure a set of pre-defined features, i.e. the width and height of each nostril in terms of pixel distance. As we will discuss in more detail in our methodology section, the measurements obtained by our respiratory extraction schema consequentially produce a subject-specific de-noised set of breathing signals representing a 60-s long breathing waveform that combines different features. In order to facilitate this, the Root Mean Square (RMS) [24] calculation is applied respectively to both left and right nostril feature values. Thus, for each frame, these values are averaged to one output, generating a hybrid temporal respiratory waveform. Second, we propose a motion classification approach when obtaining the respiratory waveform from any given video. Hence, we perform respiratory waveform peak and waveform bottom detection. Then, based on the detectable number of local maximums and minimums, the given waveforms are categorized into two groups, i.e., a nostril motion and a nostril non-motion group. In order to further reduce the number of potential candidates to match with, we define and extract six features from each input waveform, and employ a modified K-Nearest Neighbor

[25] algorithm to find the top five most similar subjects within each classified group. The key characteristics of the proposed pre-processing procedure are the following: (1) instead of comparing all subjects in the testing session, while conducting human recognition, we group the subjects in terms of their physiological responses (breath waveforms) in advance and ensure that the computational complexity of the recognition algorithm is relatively low; (2) the proposed pre-processing approach achieves very high classification accuracy (see experimental results section).

Moreover, the images employed during the face recognition process are six different combinations of facial components acquired randomly from videos under two collection scenarios: [Scenario One] - Un-controlled Condition of variant Luminance, and non-fixed head pose; [Scenario Two]—Semi-controlled condition with a fixed head position. The proposed recognition system was tested under three experiments (i.e., original images, motion classification and feature clustering, followed by motion classification, using either original or wavelet—based normalized (WA) video [26] datasets. Standard FR academic algorithms are used including, Local Binary Patterns (LBP) [27] and Local Ternary Patterns (LTP) [28]. Our proposed FR approach achieves improved rank-1 identification rates compared to the benchmark ones.

Paper Organization

The rest of this paper is organized as follows. Section "Experimental Setup" describes the experimental devices, the video datasets and the facial components used. Section "Methodology" provides a summary of the proposed respiratory wavelet extraction method, the designed pre-processing techniques, the face normalization technique used, and the face recognition algorithms employed. Section "Results and Discussion" introduces the detailed experiments conducted and the results, while conclusions as well as the future work are presented in section "Conclusions and Future Work".

Experimental Setup

Two unique facial video databases, i.e. DB_FF (fully frontal face pose) and DB HU (head-up face pose) were generated and used in the proposed study. Each database consists of two sets of video recordings (session 1 and session 2) of the same 30 subjects (data were collected on different days). For each set, five 1-min long videos were recorded for each subject (for each session, a total of 150 videos are collected). From each video database and for the purpose of conducting the facial recognition experiments, two sets of six image databases are generated (DB FF[1–6] and DB HU[1–6]) that include different facial component combinations, namely focusing on regions such as the eye, eyebrow, mouth, and their combination with the nose region (see Table 3.2). The following subsections provide more details on the data collection process.

Equipment

- **Cannon 5D Mark II**: This digital SLR camera has a 21.1-megapixel full-frame CMOS sensor with DIGIC 4 Image Processor and a vast ISO range of 100–6400. It supports Live View HD video recording with up to 39 frames per second (FPS). In this study, the camera is used to obtain ultra-high resolution facial videos in the visible spectrum.
- **ADInstruments Respiratory Belt Transducer**: This Respiratory Belt Transducer contains a piezo-electric device that measures changes in thoracic or abdominal circumference during respiration, which indicate inhalation, expiration and breathing strength and can be used to derive breathing rate. The Respiratory Belt Transducer is used to track the up-and-down movement of the chest and convert it to the breathing waveform served as our ground truth data, which helps assess our developed respiration wavelet extraction system.
- **ADInstruments PowerLab 4/35**: This PowerLab device is a high-performance data acquisition hardware, physically connected with a Respiratory Belt Transducer. It has 4 analog input channels and is capable of recording at a speed of up to 400,000 samples per second. In this study, we set up the sampling rate to be 40 samples per second, which is compatible to our video recording frame rate (39 frames per second).

Databases

Thirty (30) subjects were involved into the data collection over two sessions conducted on two different days. The following is a short description of each database utilized in our experiments. Additional information may be found in Table 3.1.

- DB_FF: Collected in a controlled indoor (CI) environment, and high quality of one 10-s video was captured for each subject for each session with the Cannon 5D Mark camera, at 39 frames per second and at 1920 × 1080 resolution. Full frontal pose face videos were collected.
- DB_HU: Collected in a controlled indoor (CI) environment. Five 60-s long videos were captured for each subject per session. We used the Cannon 5D Mark camera, at 39 frames per second and at 1920 × 1080 resolution. Videos were acquired when the pose (pitch) (see Fig. 3.2) was between 30–40 (+/−) degrees (head looking up).

Table 3.1 Utilization information about each database

Database	Cameras	# of Subjects	# of Sessions	# of Data (/subject/session)	Head pose
DB_FF	5D Mark II	30	2	1	Full frontal face
DB_HU	5D Mark II	30	2	1	Head up at 30°–40°
DB_FF1~DB_FF6	5D Mark II	30	2	1	Full frontal face
DB_HU1~DB_HU6	5D Mark II	30	2	1	Head up at 30°–40°

Table 3.2 Facial components combinations of 6 different types and their corresponding covered regions employed in the classification stage, which is to eliminate the influence of other features not contained within the specific combination

Database	Facial components	Image size	Covered area
DB_FF1~DB_HU1	Nose	180 × 150	Top_Width:55; Top_Height:40; Bottom:17
DB_FF2~DB_HU2	Nose, eye	330 × 180	Top_Width:0; Top_Height:0; Bottom:17
DB_FF3~DB_HU3	Nose, mouth	330 × 260	Top_Width:135; Top_Height:42; Bottom:22
DB_FF4~DB_HU4	Nose, eye, eyebrow	330 × 250	Top_Width:165; Top_Height:20; Bottom:22
DB_FF5~DB_HU5	Nose, mouth, eye	330 × 340	Top_Width:165; Top_Height:50; Bottom:25
DB_FF6~DB_HU6	Nose, mouth, eye, eyebrow	330 × 340	Top_Width:165; Top_Height:0; Bottom:25

- DB_F1~DB_F6: For each session, visible full frontal face images were extracted (per subject). The number 1 to 6 represents a specific image database, i.e., one of the six different facial components combinations shown in Table 3.1.
- DB_U1~DB_U6: For each session, visible (head-up face pose) face images for each subject were extracted from the corresponding group: DB U(x), where x represents a specific image database, one of the six different facial components combinations shown in Table 3.2.

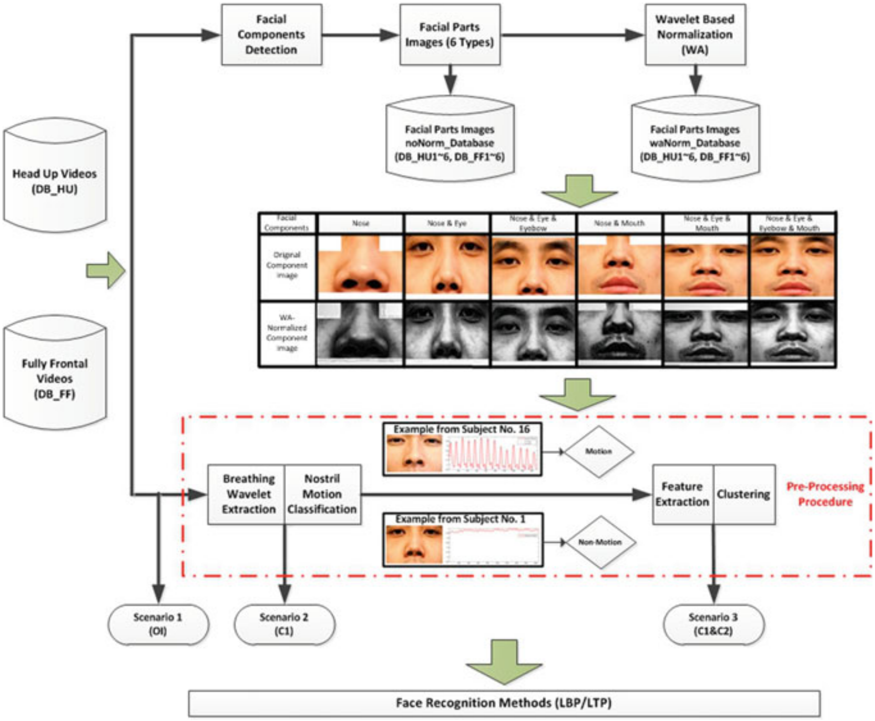

Facial Components Detection
Facial Parts Images (6 Types)
Wavelet Based Normalization (WA)
Facial Parts Images noNorm_Database (DB_HU1~6, DB_FF1~6)
Facial Parts Images waNorm_Database (DB_HU1~6, DB_FF1~6)
Head Up Videos (DB_HU)
Fully Frontal Videos (DB_FF)
Facial Components
Nose
Nose & Eye
Nose & Eye & Eyebow
Nose & Mouth
Nose & Eye & Mouth
Nose & Eye & Eyebow & Mouth
Original Component image
WA-Normalized Component image
Example from Subject No. 16
Motion
Breathing Wavelet Extraction
Nostril Motion Classification
Feature Extraction
Clustering
Pre-Processing Procedure
Example from Subject No. 1
Non-Motion
Scenario 1 (OI)
Scenario 2 (C1)
Scenario 3 (C1&C2)
Face Recognition Methods (LBP/LTP)

Fig. 3.1 Overview of the design of our proposed respiratory wavelet pre-processing approach that supports our human identity authentication system

Methodology

In this section, we outline the techniques used for data pre-processing (i.e., respiratory wavelet extraction, motion classification and feature vector clustering) and facial recognition. The overall process is displayed in Fig. 3.1 and the salient stages of the proposed approaches are described below.

Breathing Wavelet Extraction

In this study, a novel approach is proposed based on nostril movements. It uses as an input head-up facial videos (DB HU), i.e., non-contact measurements of human respiratory waveforms. During data collection, while asking subjects to breathe, we noticed that about half of them demonstrated a clear transition from normal to flared nostrils during inspiration and vice versa. This is not considered to be a medical symptom named Nasal Flaring [30], (a condition where human nostrils are dilated,

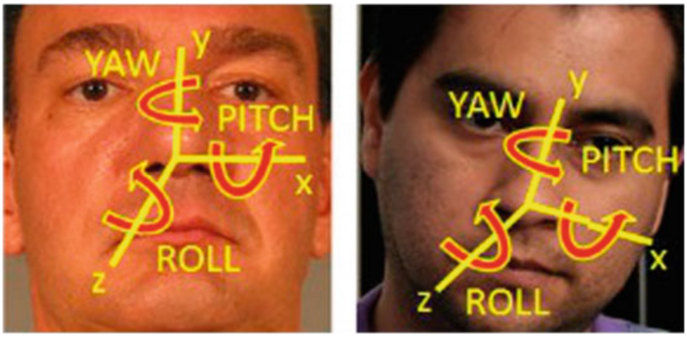


Fig. 3.2 Illustration of three types of rotational descriptors on face images acquired under variable conditions using two different sensors [29]. Note that in order to clearly track and record the movement of nostrils, the face videos (DB_UH) used in our study were generated when the pitch descriptor was measured to be between 30 and 40° plus, i.e., in head looking up position

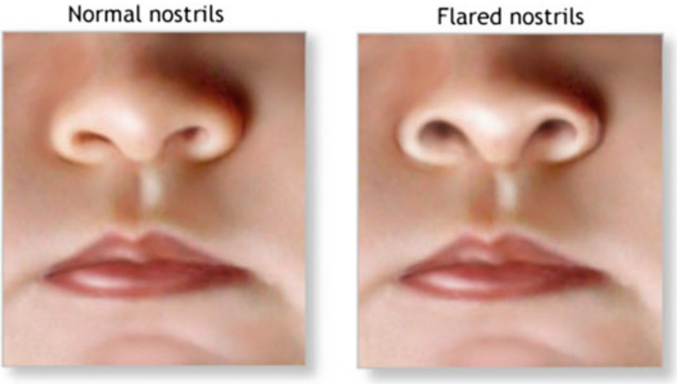


Fig. 3.3 A visual example where flared nostrils are observed. This condition is noticed over time on some individuals, and if exploited properly, we can group subjects before performing any face component-based matching (the original figure can be found at [32])

which usually occurs during inspiration and may occasionally happen during expiration or through-out the breathing cycle [31]). This was just an observation that this group of subjects was unintentionally breathing in such a way. In the other group of subjects, the normal to flaring transition behavior was not that obvious but it was a significantly enough signal to be exploited for the purpose of this study (Fig. 3.3).

The proposed recognition method, firstly, detects the nose position using the Masayuki Tanaka face parts detection algorithm [33] for each frame in a target video. Then, it measures the horizontal width W and the vertical height H for both nostrils (i.e., $H_{left}, W_{left}, H_{right}, W_{right}$). Figure 3.4 illustrates the length measurement approach for one frame. By combining each type of distance collected from any single video, we managed to obtain 4 sets of data (as discussed above) representing the

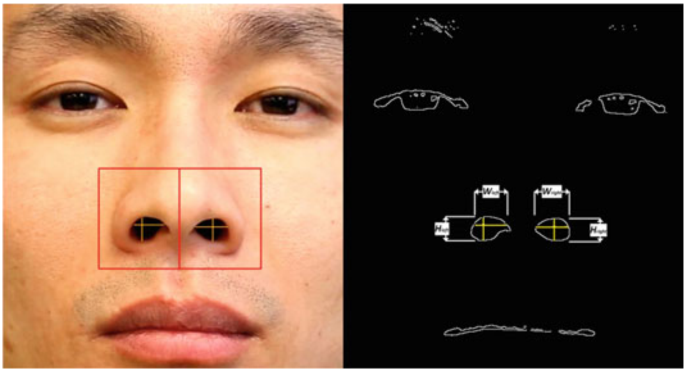


Fig. 3.4 Illustration of our nostrils features measurement approach. The nose location is first detected (shown as red squares) using the Masayuki Tanaka face parts detection algorithm [33]. Then, 4 features (left height H_{left}, left width W_{left}, right height H_{right} and right width W_{right}) are measured. This face is selected from a random subject and at a random video of the data collection (Section #1; Video #1; Frame #12)

nostrils movement within 1-min time frame. Considering the measurement errors (such as the slight distance difference between the camera and the subjects during the video shooting process), all 4 sets of distances were processed via a 0–1 normalization process (scaling all values between 0 and 1), using formula (1) and applying a signal de-noising algorithm. Figure 3.5 shows the breathing wavelet extraction process of converting the measured nostril features from each frame into curves.

$$Normalized(e_i) = \frac{e_i - S_{min}}{S_{max} - S_{min}} \tag{1}$$

where e_i is the target value to be normalized; S_{max} and S_{min} are the maximum and minimum values in the given set of data. Instead of analyzing nostril features separately, the Root Mean Square [24] (RMS, also called quadratic mean) measurement is computed to combine all feature values (i.e., H_{left}, W_{left}, H_{right}, and W_{right}) collected from one given frame and extract one mean target. The RMS is mathematically characterized as:

$$L_{left} = \sqrt{\frac{{H_{left}}^2 + {W_{left}}^2}{2}} \tag{2}$$

$$L_{right} = \sqrt{\frac{{H_{right}}^2 + {W_{right}}^2}{2}} \tag{3}$$

$$L_{mean} = \frac{L_{left} + L_{right}}{2} \tag{4}$$

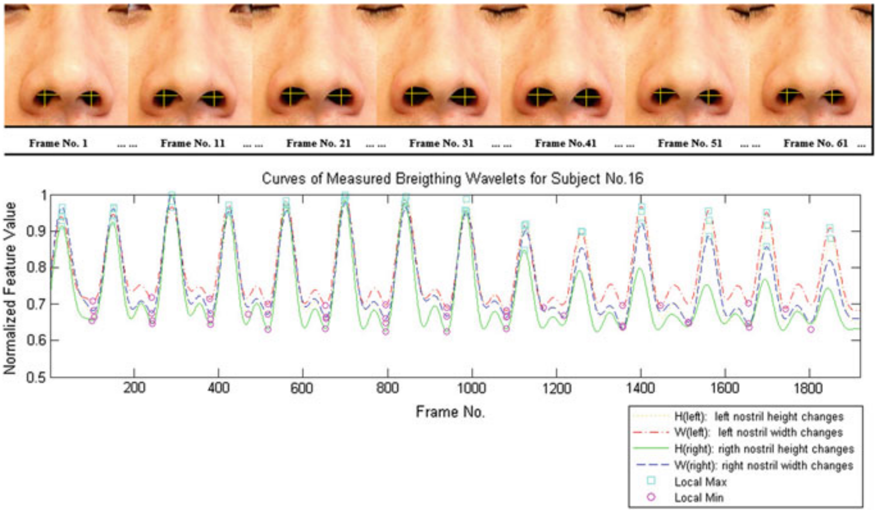


Fig. 3.5 Example of extracting breathing wavelets from a given video. The proposed approach keeps tracking and measuring 4 nostril features of the subject in each frame of the video. Then apply 0–1 normalization and de-noising algorithm to each collected data set (a data set includes all the measured values for a specific nostril feature in a video). (Top) The pictures are selected with 10-frame interval to display the observable nostril movement. (Bottom) It is the resulting 4 curves for the measured nostril features within one video. Both the frames and the curves are gathered from subject no. 16 (section no. 1, video no. 1)

where H_{left}, W_{left}, H_{right}, and W_{right} are the measured nostril feature values from the given frame; L_{left} and L_{right} are the resulting left and right RMS values and L_{mean} is the target mean value combining 4 features. The resulting breathing wavelet for our designed system is generated according to the collected set of the mean value L_{mean} based on left RMS value L_{left} and right RMS value L_{right} for each frame. The comparison between the output waveform extracted from one sample video and its corresponding ground truth breathing curve (i.e. real-time wavelet detected via ADInstruments device) is shown in Fig. 3.6 above.

Nostril Motion Classification

After applying breathing wavelet extraction to the videos, 5 new data sets are generated for each subject in both sessions. Each data set can be represented as a 60-s long respiratory waveform extracted from head-up facial videos in database DB_HU. By roughly comparing the waveforms, we observed that some subjects show comparatively high nostril movement (large amplitude in waveform), while he same movement in other subjects is less noticeable (small amplitude in waveform). Our proposed nostril motion classification technique is based on the detectable number of

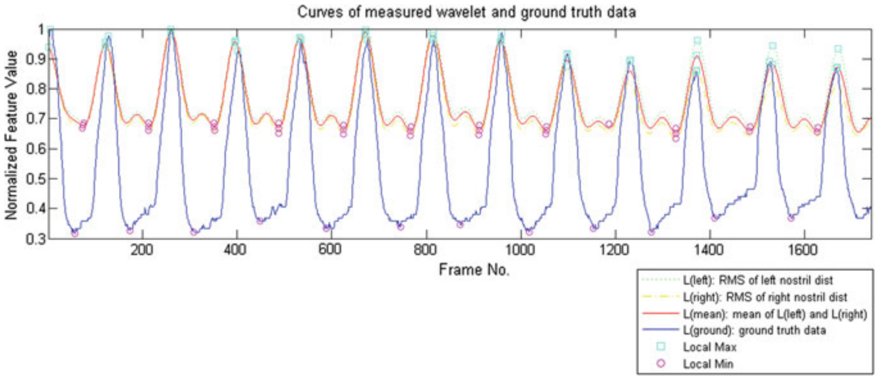


Fig. 3.6 Comparison between the breathing wavelet output from our designed system (red line) and the ground truth breathing curve measured by ADInstruments device (blue line). The waveforms are selected from subject no. 16 (section no. 1, video no. 1) which is the same subject shown in Figs. 3.4 and 3.5

both peak points and bottom spots of a given breathing wavelet using the Peakdet algorithm [33] with fixed threshold values (i.e., the pre-define value used to determine the local maximum and local minimum point, the thresholds of detected number of peaks and bottoms to classify movement). Figure 3.7 illustrates the nostril movement classification process. The defined classifier worked on both sessions and categorized subjects of each session into two classes: Movement Group (MG) and Non-Movement Group (NG). Since each subject has 5 distinct data sets on each unit (i.e., unit is defined as a group of data for one subject in one session), the motion classification output is referred to the majority voting results within the specific unit. The detailed decision making procedure for our motion classifier is displayed in Table 3.3.

Feature Extraction and Clustering

In order to further narrow down the number of potential candidates as well as reduce the matching complexity in the human recognition stage, the clustering technique based on respiratory wavelet features was introduced in the pre-processing system. The following content provides more details about the feature definition and clustering procedure.

Feature Extraction

Inspired by the studies of human electrocardiogram (ECG) identification, we proposed the hypothesis that some crucial features can be extracted from breathing

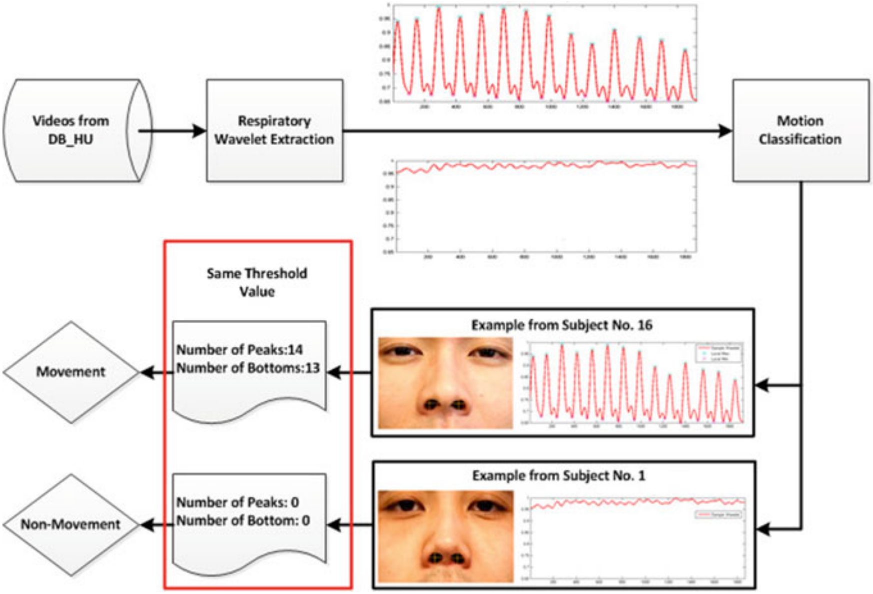


Fig. 3.7 Illustration of how our nostril motion classifier that categorizes input subjects into movement and non-movement groups. 2 examples are selected from subject #16 (section #1, video #1) and subject #1 (section #1, video #3). The threshold value for detecting peak and bottom point is 0.15 and the number of local maximums (as well as minimums) used for classification is 4

Table 3.3 The detailed decision making procedure for our motion classifier

Decision	Decision	
5	0	Motion
4	1	Motion
3	2	Motion
2	3	Non-motion
1	4	Non-motion
0	5	Non-motion

wavelets to help categorize subjects. Referred to the defined feature set used for ECG classification [10, 34–36], 4 dynamic features are selected within each breathing cycle time (i.e., one inspiration and its following expiration) shown in Fig. 3.8.

To represent a 60-s long respiratory wavelet, instead of using feature values extracted from one periodic time, our defined feature consists of averaged attribute values among the detected integrated cycles. In addition, 2 more features are defined as the ratio of horizontal and vertical distances for both left and right nostrils. They are extracted from facial images to complement the feature vector representing one breathing wavelet. The detailed definition about each attribute is provided in Table 3.4.

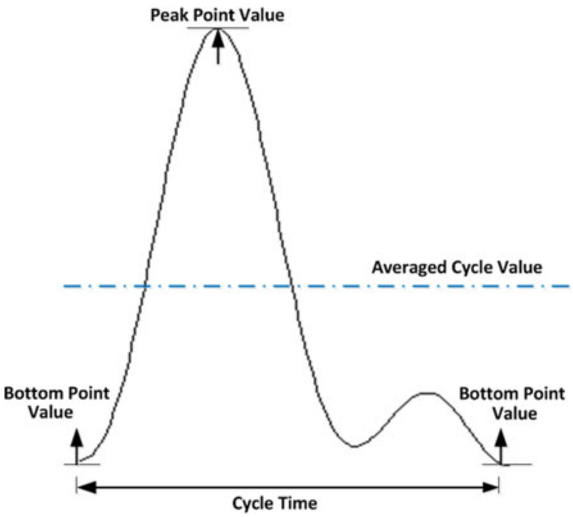


Fig. 3.8 Definition of 4 dynamic features for a cycle of breathing wavelet. Notice that its a real respiratory cycle extracted from the output of our breathing waveform extraction system for subject #16

Table 3.4 Description of the features defined on each input video and its corresponding respiratory wavelet which consist the feature vector for future subjects clustering. Notice that all the features are defined and extracted from one generated respiratory wavelet

Feature	Feature definition
F1	Ratio of horizontal distance and vertical distance for left nostril
F2	Ratio of horizontal distance and vertical distance for right nostril
F3	Averaged cycle time for one inspiratory and its expiratory in 60-s
F4	Averaged peak value in 60-s
F5	Averaged bottom value in 60-s
F6	Averaged value for one wavelet in 60-s

Clustering

After feature extraction, for each subject 5 distinct feature vectors are generated for each session. Then the modified K-Nearest Neighbor (KNN) [25] approach is used to further cluster subjects after nostril motion classification. The designed KNN model is to treat five feature vectors for every subject in each session as a group. In terms of each group in training session (i.e., session 1) the cluster will find the top 5 groups with the smallest averaged distance of the given input group in testing session (i.e., session #2). The rules of distance measurement between 2 given groups is defined as: Step (1) Each vector from one group is used to compute the Euclidean Distance to all vectors from the other group; Step (2) We then average all 25 distance values produced from Step (1); (3) Set he averaged value as the distance between these two given sets. Mathematically this is described as:

$$dist(v_i, v_j) = \sum_{k=1}^{6}(f_{ik} - f_{jk})^2 \tag{5}$$

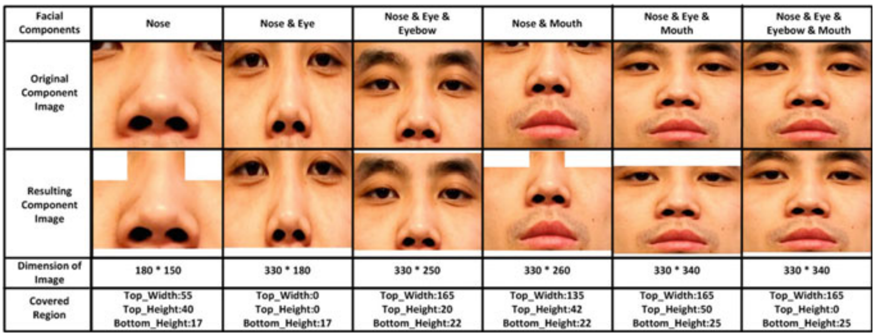
$$dist_value = \frac{1}{25}\sum_{i=1}^{5}\sum_{j=1}^{5} dist(v_i, v_j) \tag{6}$$

where v_i and v_j (i = $1, 2, \ldots, 5$ and j = $1, 2, \ldots, 5$) are the feature vectors that come from the compared training group and testing group, and f_{ik} (k = $1, 2, \ldots, 6$) is the corresponding kth feature value for specific vector v_i.

Component-Based Face Recognition

The major steps of the recognition process are described below:

- **Face Image Selection**: 5 frames of facial images are randomly selected from each frontal face video (DB_FF) as well as the head-up pose video (DB_HU) for every subject in both sessions.
- **Image Reconstruction**: Since subjects may change their head pose during the video taking process, an image modification scheme was applied to all face images. The approach consists of three main steps: pupil detection, image rotation and image resizing. Automated eye detection was performed using the Masayuki Tanaka face parts detection algorithm [33], where the coordinates of the pupils were automatically obtained. After the eye centers were located, the canonical faces were automatically constructed by employing an affine transformation to achieve horizontally rearranged eye locations. Finally, all facial images were resized to obtain fixed distance between two pupils (i.e., 180 pixels).
- **Facial Components Extraction**: Our face recognition databases of both the frontal face pose and the head-up pose (DB_FF1∼DB_FF6 and DB_HU1∼ DB_HU6) are consisted of 6 diverse combinations of facial parts extracted from all selected images after reconstruction, and the details about each type of combination is shown in Table 3.2. After acquiring the pupil coordinates, the facial components segmentation was automatically conducted with the pre-defined cutting locations for each combination. Since the proportion of facial parts may vary among different faces, the resulting facial part segmentations may contain other features, such as the cutting of the nose component may include partial canthus structure. Therefore, for each type of facial components segmentation image a fixed cover strategy was applied. Table 3.2 provides more details about the predetermined covered regions for each type of facial parts combination. Figure 3.9 displays an example of 6 types of facial components combinations of the original and the resulting segmentation image with specific covered area.



Facial Components	Nose	Nose & Eye	Nose & Eye & Eyebow	Nose & Mouth	Nose & Eye & Mouth	Nose & Eye & Eyebow & Mouth
Original Component Image						
Resulting Component Image						
Dimension of Image	180 * 150	330 * 180	330 * 250	330 * 260	330 * 340	330 * 340
Covered Region	Top_Width:55 Top_Height:40 Bottom_Height:17	Top_Width:0 Top_Height:0 Bottom_Height:17	Top_Width:165 Top_Height:20 Bottom_Height:22	Top_Width:135 Top_Height:42 Bottom_Height:22	Top_Width:165 Top_Height:50 Bottom_Height:25	Top_Width:165 Top_Height:0 Bottom_Height:25

Fig. 3.9 An example (subject no. 1) of the comparison between the original facial components combinations and the resulting segmentation images with pre-defined covered region for each type of combination. (In order to show 6 different types of facial parts combination, all the images were resized to the same size.)

- **Image Normalization**: 14 types of normalization approaches (i.e., WA, AS, DCT, GRA, HOMO, IS, LSSF, MAS, MSR, MSW, SF, SSR, TT and WEB) provided in Illumination Normalization techniques for robust Face recognition (INface) toolbox v2.0 [37, 38] have been tested in our study, and Fig. 3.10 compares all kinds of resulting normalized images for a given sample input. More detailed description about each normalization method is introduced in Appendix A. Based on experimental output, the wavelet based normalization technique (WA) provided the best experimental results and was applied to all facial components combination images in DB_FF1~DB_FF6 and DB_HU1~DB_HU6. The WA approach was proposed by Du and Ward [26] that applies the discrete wavelet transform to an image and processes the obtained sub-bands. This technique focuses on the detailed coefficient matrices and employs histogram equalization to the approximate transform coefficients. Finally, our proposed approach reconstructs the normalized image using the inverse wavelet transform after the manipulation of each individual subband.
- **Facial Parts Recognition Method**: The standard texture based face recognition method was employed to perform the facial components recognition experiments, such as the Local Binary Patterns (LBP) [27, 39] and the Local Ternary Patterns (LTP) [28]. The LBP works as a gray-scale invariant texture measure, derived from a general definition of texture in a local neighborhood. The LBP operator labels the image pixels by setting the threshold of a 3 by 3 neighborhood based on the value of the center pixel and the resulting binary pattern is converted to a decimal value used for labeling the given pixel. A local neighborhood is defined as a set of sampling points evenly spaced on a circle. The LTP operator is introduced as an extension of LBP. The identification performance is evaluated through the Calculative Match Characteristic (CMC) curve, which measures the 1:m recognition performance and evaluates the ranking capability of the system [40].

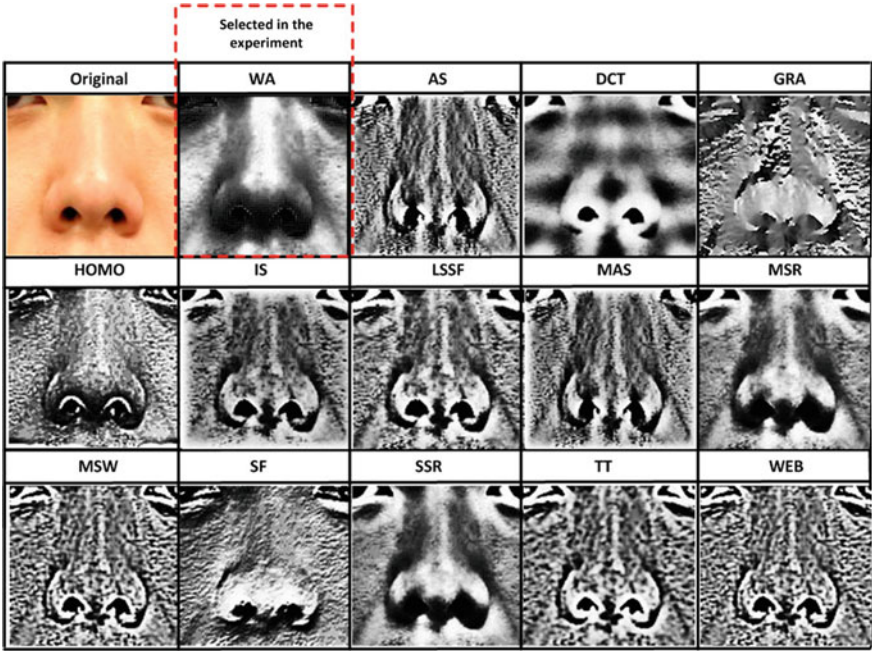


Fig. 3.10 An example (subject no. 1) of the comparison between the original facial components combinations and the resulting segmentation images with pre-defined covered region for each type of combination. (In order to show 6 different types of facial parts combination, all the images were resized to the same size.)

Results and Discussion

The face recognition experiments (please see Table 3.5 below) worked on the databases of all facial parts combinations (i.e., DB_HU1~DB_HU6 and DB_FF1~DB_FF6) for both WA-normalized and non-normalized images under the following scenarios:

- **Original Images (OI)**: Face recognition worked on both WA-normalized and non-normalized facial parts combination images.
- **Motion Classification (C1)**: Firstly, subjects are categorized into two groups, then face recognition is applied separately on each group of both WA-normalized and original facial parts combination images.
- **Motion Classification and Feature Clustering (C1&C2)**: Firstly, subjects are categorized into two groups. Then, for each subject, the top 5 most similar entities via feature clustering are found, before, finally face recognition is applied on each cluster of the 5 candidates in both WA-normalized and non-normalized form of facial components combination images.

Motion Classification Results (C1)

For all waveforms generated by applying the respiratory wavelet extraction system to every video in DB_HU, the motion classifier was applied. The classifier can categorize the input waveforms produced by identical subjects in different sessions into same group (Motion Group or Non-Motion Group) with accuracy of 100%. The detailed classification output for each subject can be found in Table 3.6.

Feature Clustering Results (C1&C2)

After categorizing subjects into 2 groups, extracted feature vectors for each waveform was used to manipulate further clustering. As a result, for each subject in the

Table 3.5 Rank 1 identification rates when utilizing CMC curves (LBP) for OI, C1, and C1&C2 scenarios of both WA-normalized (waNorm) and Non-normalized (noNorm) matching

DB_HU

	Scenario	DB_HU1	DB_HU2	DB_HU3	DB_HU4	DB_HU5	DB_HU6
noNorm	OI	0.80	0.93	0.89	0.95	0.91	0.92
	C1	0.84	0.95	0.90	0.95	0.91	0.93
	C1&C2	0.87	0.95	0.89	0.97	0.92	0.94
waNorm	OI	0.82	0.92	0.89	0.99	0.92	0.96
	C1	0.85	0.93	0.89	0.99	0.93	0.96
	C1&C2	0.89	0.94	0.90	0.97	0.95	0.95

DB_FF

	Scenario	DB_FF1	DB_FF2	DB_FF3	DB_FF4	DB_FF5	DB_FF6
noNorm	OI	0.93	0.96	0.97	0.97	0.97	0.97
	C1	0.96	0.98	0.97	0.97	0.97	0.97
	C1&C2	0.93	0.95	0.93	0.94	0.93	0.93
waNorm	OI	0.89	0.97	0.96	0.99	0.99	1
	C1	0.95	0.98	0.97	0.99	0.99	1
	C1&C2	0.93	0.95	0.97	0.97	0.96	0.97

Table 3.6 Output of two categorizations (movement and non-movement group) from nostril movement classification for each subject in both sessions of database DB_HU

Group	# of Subjects	Subjects' ID
Motion Group (MG)	14	6, 7, 8, 11, 12, 14, 16, 17, 18, 19, 22, 23, 24, 29
Non-Motion Group (NG)	16	1, 2, 3, 4, 5, 9, 10, 13, 15, 20, 21, 25, 26, 27, 28, 30

Table 3.7 Accuracies of finding the identical subject in the output group after employing feature vector clustering

Group name	Accuracy(%)
Motion Group (MG)	100
Non-Motion Group (NG)	93.75
All Subjects	96.67

training session, the cluster will generate an output consisting of the top 5 most similar subjects in the testing session. The accuracy of obtaining the identical subject in the output cluster is shown in Table 3.7.

Component-Based FR Experiments for Three Scenarios

The CMC performance (LBP approach) for scenarios 1–3 across all databases (DB_HU1~DB_HU6 and DB_FF1~DB_FF6) for 2 types of images (waNorm and noNorm) is illustrated in Figs. 3.11 and 3.12. In addition the performance measured by rank-1 identification rate of the proposed 3 experiments (i.e., OI, C1 and C1&C2) are presented in Table 3.5.

The experimental results indicate that human identify matching using only nose part under uncontrolled situation is a very challenging problem. Since the identification rate at rank-1 for original image (i.e., scenario OI, baseline scenario) of head-up pose is 80% for non-normalized form and 82% after applying WA normalization technique. And both gallery and probe images were randomly acquired from videos under possibly different illumination situation as well as uncontrolled head poses while taking videos. However, when using our proposed motion classification approach (i.e., scenario C1), the identification rate is increased to 84% for the original images and 85% with WA-normalization. Moreover, when the feature clustering technique is respectively applied on the two classified categorizations (i.e., scenario C1&C2), the identification rate at rank-1 improves about 7%, resulting in 87% accuracy for original input and 89% accuracy with WA-normalization manipulation compared to baseline scenario.

In the other databases (DB_HU2~DB_HU6), more facial biometric features (i.e., eye, eyebrow and mouth) are included besides nose component. Although the baseline experimental results are increased to over 90% matching accuracy at uncontrolled condition, our feature clustering approach in combination with motion classification technique can still improve the identification rate (at rank-1) by approximately 2%.

Under the semi-controlled condition (DB_FF1~DB_FF6), the head pose were fixed to fully frontal position. However the illumination status still depended on the lighting conditions when the data were collected. Compared with the baseline

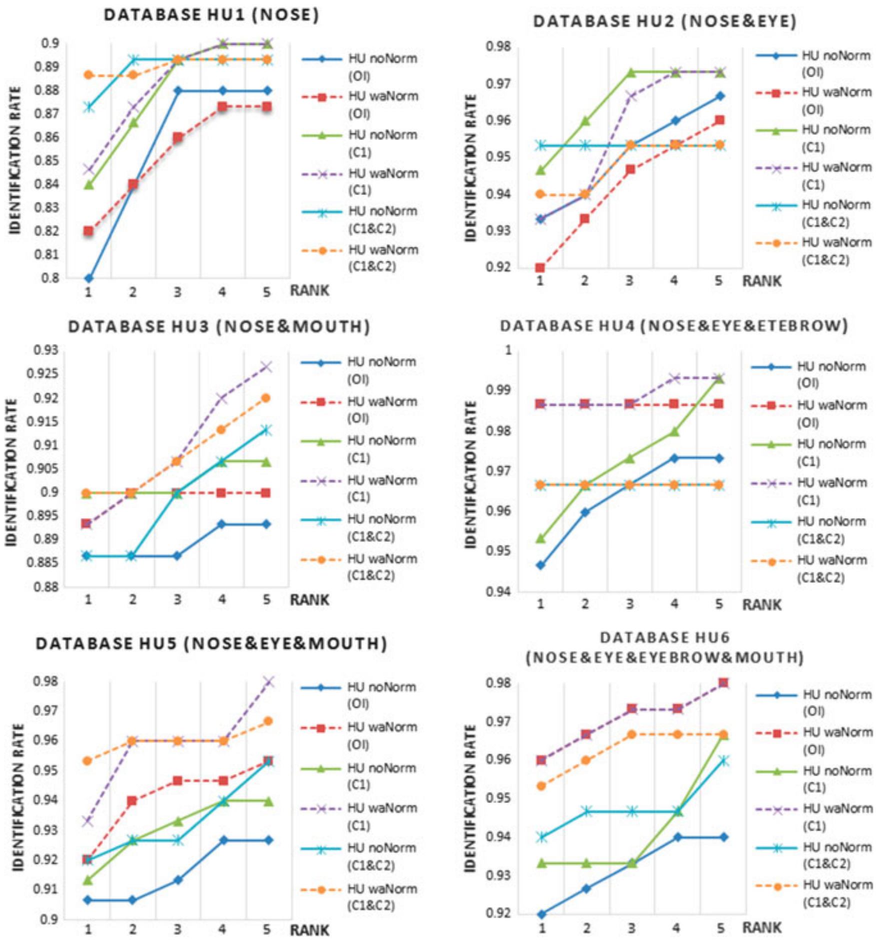


Fig. 3.11 CMC curves comparing the performance of 3 types of scenarios (i.e., OI, C1 and C1&C2) for head-up face images of 6 different facial parts combinations (DB_HU1~DB_HU6)

situation of uncontrolled condition, the identification rates of the baseline scenario with different types of facial component combinations in both WA-normalized and original forms (i.e., Scenario OI) increased. Our proposed motion classification technique can still remain good performance. Especially for database DB_FF1 with only the nose information, the advantage of using motion classification is that it performs even better than baseline performance by approximately 3% for non-normalized and 6% for WA-normalized form. However, in this semi-controlled condition, the feature clustering technique doesn't perform well. The main problem is that although the motion classification can produce 100% accuracy, the following feature clustering approach still introduces some error (i.e., overall accuracy is 97% with 100%

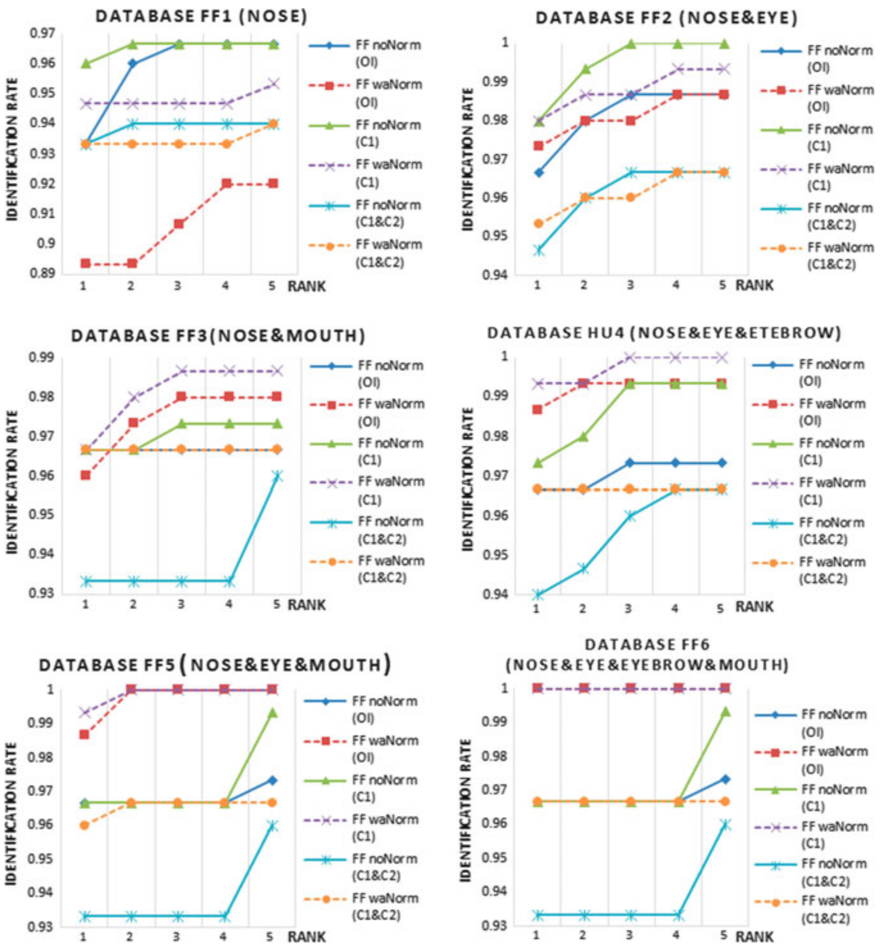


Fig. 3.12 CMC curves comparing the performance of 3 types of scenarios (i.e., OI, C1 and C1&C2) for fully frontal face images of 6 different facial parts combinations (DB_FF1~DB_FF6)

accuracy in motion group and 94% for non-motion group), which will lead to the output identification rate at best as 97%.

Conclusions and Future Work

In this paper our focus was on investigating the problem of human identity recognition in uncontrolled environments with the help of respiratory wavelets extracted from videos, when taking advantage of a medical condition known as nasal flaring. Specifically we have studied the human facial components identification tasks under several pre-processing techniques. The breathing waveform produced by our proposed respiratory wavelet extraction system is comparative to the ADInstruments device measured output for some portion of subjects showing nostril flaring symptom. The proposed motion classification technique can successfully categorize input waveforms into two groups with 100% accuracy. Notice that the defined accuracy is to classify the wavelets in both data collection sessions for the same subjects into the same group. In addition, our designed feature clustering technique, combined with our proposed motion classification approach, can locate the same subject within the top 5 candidates with an accuracy of 97% (i.e., 100% accuracy for motion group and 94% accuracy for non-motion group), which results in significant reduction on the number of comparisons during the identity recognition procedure.

For the purpose of this work, 3 different scenarios (i.e., Original Image, Motion Classification and Feature Clustering) were designed and tested on 6 types of databases (DB_HU1∼DB_HU6 and DB_FF1∼DB_FF6). And the experimental databases consist of different facial component combinations in both non-normalized form and Wavelet Based (WA) normalized form under semi-controlled as well as uncontrolled indoor environments. The experimental results indicate that, across all datasets used, the application of motion classification and feature clustering improves facial components recognition performance. Especially when only investing the nose information, our proposed system can increase the rank-1 identification rate by approximately 7% for both original and WA-normalized datasets. The proposed respiratory wavelet recognition technique shows particularly good performance on the baseline scenarios with low accuracy under uncontrolled conditions.

Our future plans are to develop an improved respiratory wavelet extraction approach that will successfully detect and enhance the slight movement of nostrils.

In addition, the experiment video databases will be acquired using cellphone cameras with diverse resolution under uncontrolled conditions, such as varying luminous condition and unfixed head position. This is expected to result in as good as or even improved identification performance.

Face Normalization Techniques

- **Wavelet Based Normalization Technique (WA)**: The WA approach was proposed by Du and Ward [26] that applies the discrete wavelet transform to an image and processes the obtained sub-bands. This technique focuses on the detailed coefficient matrices and employs histogram equalization to the approximate transform coefficients. Finally, reconstructs the normalized image using the inverse wavelet transform after the manipulation of each individual sub-band.
- **Anisotropic Diffusion Based Normalization Technique (AS)**: The AS approach adopts anisotropic smoothing of the input image to evaluate the luminance function, which was introduce by Gross et al. [41] to the face recognition area.
- **Discrete Cosine Transform Based Normalization Technique (DCT)**: The DCT technique was introduced by Chen et al. [42] that truncates an appropriate number

of DCT coefficients to minimize illumination variations under different lighting conditions.

- **Gradientfaces Based Normalization Technique (GRA)**: The GRA approach proposed by Zhang et al. [43] is to transform image into the gradient domain and use the generated face representation as the illumination invariant version of the target image.
- **Homomorphic Filtering Based Normalization Technique (HOMO)**: The HOMO is to transform the input image into the logarithm followed by the frequency domain in order to enhance the high-frequency components and weaken the low-frequency parts. Finally, apply the inverse Fourier transform to obtain the output image in the spatial domain [44].
- **Isotropic Diffusion Based Normalization Technique (IS)**: The IS approach [44] is to estimate the luminance function of the input image using isotropic smoothing algorithm which is a simpler variance of the anisotropic diffusion based normalization technique [41].
- **Large-Scale and Small-Scale Features Normalization Technique (LSSF)**: The LSSF proposed by Xie et al. [45] firstly computes the reflectance and luminance function of the input image and then further analyzes both generated functions using a second time of normalization. Within the INface toolbox used in our experiment, the SSR technique is implemented as the normalization approach in both steps of LSSF technique.
- **Modified Anisotropic Diffusion Normalization Technique (MAS)**: The MAS approach included two main modification into the original anisotropic diffusion normalization technique [41]: (1) Introducing an additional atan function to estimate the local contrast; (2) Apply a robust post-processing procedure proposed by Tan et al. [28] in the final stage of this technique.
- **Multi-Scale Retinex (MSR) Algorithm**: The MSR method is to extend the previously designed single-scale center/surround retinex technique to a multi-scale version proposed by Jobson et al. [46].
- **Multi-Scale Weberfaces Normalization Technique (MSW)**: The MSW as an extend the single-scale Weberfaces approach proposed by Wange et al. [47] is to compute the relative gradient using a modified Weber contrast method for diverse neighborhood sizes and to apply a linear combination of the produced face representation as an illumination invariant version of the target output.
- **Steerable Filter Based Normalization Technique (SF)**: The SF approach produces the target normalized image by removing illumination induced appearance variation from the input facial image using steerable filters.
- **Single Scale Retinex (SSR) Algorithm**: The SSR approach was proposed by Jobson et al. [48] on the basis of the retinex theory [49] as the majority of photometric normalization techniques.
- **Tan and Triggs Normalization Technique (TT)**: The TAT is to employ a processing chain on the input image by firstly using gamma correction, then applying DoG filtering and finally adopting a robust post-processor to generate the output normalized image [28].

- **Single Scale Weberfaces Normalization Technique (WEB)**: The WEB method is to compute the relative gradient using a modified Weber contrast algorithm and treat the generated face representation as an illumination invariant version of the target image [47].

References

1. Adinstruments. http://www.adinstruments.com/
2. Jain AK, Ross A, Pankanti S (2006) Biometrics: a tool for information security. IEEE Trans Inf Forensics Secur 1(2):125–143
3. Jain AK, Bolle R, Pankanti S (1999) Biometrics: personal identification in networked society. Springer
4. Maltoni D, Maio D, Jain AK, Prabhakar S (2009) Handbook of fingerprint recognition. Springer
5. Jain AK, Li SZ (2005) Handbook of face recognition. Springer
6. Sanchez-Reillo R, Sanchez-Avila C, Gonzalez-Marcos A (2000) Biometric identification through hand geometry measurements. IEEE Trans Pattern Anal Mach Intell 22(10):1168–1171
7. Daugman J (2004) How iris recognition works. IEEE Trans Circuits Syst Video Technol 14(1):21–30
8. Srinivasan D, Ng W, Liew A (2006) Neural-network-based signature recognition for harmonic source identification. IEEE Trans Power Deliv 21(1):398–405
9. Liu S, Silverman M (2001) A practical guide to biometric security technology. IT Prof 3(1): 27–32
10. Kyoso M, Uchiyama A (2001) Development of an ECG identification system. In: Proceedings of the 23rd annual international conference of the IEEE engineering in medicine and biology society, 2001, vol 4. IEEE, pp 3721–3723
11. Irvine J, Wiederhold B, Gavshon L, Israel S, McGehee S, Meyer R, Wiederhold M (2001) Heart rate variability: a new biometric for human identification. In: Proceedings of the international conference on artificial intelligence (IC-AI01), pp 1106–1111
12. Israel SA, Scruggs WT, Worek WJ, Irvine JM (2003) Fusing face and ECG for personal identification. In: 2003 Proceedings of the 32nd applied imagery pattern recognition workshop. IEEE, pp 226–231
13. Israel SA, Irvine JM, Cheng A, Wiederhold MD, Wiederhold BK (2005) ECG to identify individuals. Pattern Recogn 38(1):133–142
14. Travaglini A, Lamberti C, DeBie J, Ferri M (1998) Respiratory signal derived from eight-lead ECG. In: Computers in cardiology 1998. IEEE, pp 65–68
15. Zhang T, Keller H, OBrien MJ, Mackie TR, Paliwal B (2003) Application of the spirometer in respiratory gated radiotherapy. Med Phys 30(12):3165–3171
16. Marks MK, South M, Carter BG (1995) Measurement of respiratory rate and timing using a nasal thermocouple. J Clin Monit 11(3):159–164
17. Allison RD, Holmes E, Nyboer J (1964) Volumetric dynamics of respiration as measured by electrical impedance plethysmography. J Appl Physiol 19(1):166–173
18. Heisele B, Ho P, Poggio T (2001) Face recognition with support vector machines: global versus component-based approach. In: 2001 Proceedings of the eighth IEEE international conference on computer vision, ICCV 2001, vol 2. IEEE, pp 688–694
19. Heisele B, Koshizen T (2004) Components for face recognition. In: 2004 Proceedings of the sixth IEEE international conference on automatic face and gesture recognition. IEEE, pp 153–158
20. Huang J, Heisele B, Blanz V (2003) Component-based face recognition with 3D morphable models. In: Audio-and video-based biometric person authentication. Springer, pp 27–34

21. Heisele B, Serre T, Pontil M, Vetter T, Poggio T (2001) Categorization by learning and combining object parts. In: NIPS, pp 1239–1245
22. Osia N, Bourlai T (2014) A spectral independent approach for physiological and geometric based face recognition in the visible, middle-wave and long-wave infrared bands. Image Vis Comput (in press)
23. Osia N, Bourlai T (2012) Holistic and partial face recognition in the MWIR band using manual and automatic detection of face-based features. In: 2012 IEEE conference on technologies for homeland security (HST). IEEE, pp 273–279
24. Kenney JF, Keeping ES (1962) Root mean square. In: Mathematics of statistics, 3rd edn. Princeton, NJ, Van Nostrand, pp 59–60
25. Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 46(3):175–185
26. Du S, Ward R (2005) Wavelet-based illumination normalization for face recognition. In: 2005 IEEE international conference on image processing. ICIP 2005, vol 2. IEEE, pp II–954
27. Ahonen T, Hadid A, Pietikäinen M (2004) Face recognition with local binary patterns. In: Computer vision-ECCV 2004. Springer, pp 469–481
28. Tan X, Triggs B (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans Image Process 19(6):1635–1650
29. Bourlai T, Whitelam C, Kakadiaris I (2011) Pupil detection under lighting and pose variations in the visible and active infrared bands. In: 2011 IEEE international workshop on information forensics and security (WIFS). IEEE, pp 1–6
30. Kaneshiro NK, Zieve D (2012) Nasal flaring. http://www.nlm.nih.gov/medlineplus/ency/article/003055.htm (MedlinePlus)
31. Wilkins LW (2007) Sensory system. In: Lippincott manual of nursing practice series: assessment. Lippincott Williams & Wilkins, pp 265–266
32. Kaneshiro NK, Zieve D. Nasal flaring, MedlinePlus. http://www.nlm.nih.gov/medlineplus/ency/imagepages/17279.htm
33. Tanaka M (2012) Face parts detection algorithm. http://www.mathworks.com/matlabcentral/fileexchange/36855-face-parts-detection (updated in 2014)
34. Biel L, Pettersson O, Philipson L, Wide P (2001) ECG analysis: a new approach in human identification. IEEE Trans Instrum Meas 50(3):808–812
35. Wang Y, Agrafioti F, Hatzinakos D, Plataniotis KN (2008) Analysis of human electrocardiogram for biometric recognition. EURASIP J Adv Signal Process 2008:19
36. Agrafioti F, Gao J, Hatzinakos D (2011) Heart biometrics: theory, methods and applications. In: Biometrics: book, vol 3, pp 199–216
37. Štruc V, Pavešić N (2009) Gabor-based kernel partial-least-squares discrimination features for face recognition. Informatica 20(1):115–138
38. Štruc V, Pavešic N (2011) Photometric normalization techniques for illumination invariance. Adv Face Image Anal Tech Technol IGI Global 279–300
39. Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: application to face recognition. IEEE Trans Pattern Anal Mach Intell 28(12):2037–2041
40. Kalka ND, Bourlai T, Cukic B, Hornak L (2011) Cross-spectral face recognition in heterogeneous environments: a case study on matching visible to short-wave infrared imagery. In: 2011 International joint conference on biometrics (IJCB). IEEE, pp 1–8
41. Gross R, Brajovic V (2003) An image preprocessing algorithm for illumination invariant face recognition. In: Audio-and video-based biometric person authentication. Springer, pp 10–18
42. Chen W, Er MJ, Wu S (2006) Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. IEEE Trans Syst Man Cybern Part B Cybern 36(2):458–466
43. Zhang T, Tang YY, Fang B, Shang Z, Liu X (2009) Face recognition under varying illumination using gradientfaces. IEEE Trans Image Process 18(11):2599–2606
44. Heusch G, Cardinaux F, Marcel S (2005) Lighting normalization algorithms for face verification. In: IDIAP-com 05-3

45. Xie X, Zheng W-S, Lai J, Yuen PC, Suen CY (2011) Normalization of face illumination based on large-and small-scale features. IEEE Trans Image Process 20(7):1807–1821
46. Jobson DJ, Rahman Z-U, Woodell GA (1997) A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Trans Image Process 6(7):965–976
47. Wang B, Li W, Yang W, Liao Q (2011) Illumination normalization based on Weber's law with application to face recognition. IEEE Signal Process Lett 18(8):462–465
48. Jobson DJ, Rahman Z-U, Woodell GA (1997) Properties and performance of a center/surround retinex. IEEE Trans Image Process 6(3):451–462
49. Land EH, McCann J (1971) Lightness and retinex theory. JOSA 61(1):1–11

Chapter 4
A Study on Human Recognition Using Auricle and Side View Face Images

Susan El-Naggar, Ayman Abaza and Thirimachos Bourlai

Abstract Face profile, the side view of the face, provides biometric discriminative information complimentary to the information provided by frontal view face images. Biometric systems that deal with non-cooperative individuals in unconstrained environments, such as those encountered in surveillance applications, can benefit from profile face images. Part of a profile face image is the human ear, which is referred to as the auricle. Human ears have discriminative information across individuals and thus, are useful for human recognition. In the current literature, there is no clear definition for what a face profile is. In this study, we discuss challenges related to this problem from recognition performance aspect to identify which parts of the head side view provide distinctive identity cues. We perform an evaluation study assessing the recognition performance of the distinct parts of the head side view using four databases (FERET, WVU, UND, and USTB). The contributions of this paper are three-fold: (i) by investigating which parts of the head side view increase the probability of successful human authentication, we determined that ears provide main features in the head side view. The rank-1 identification performance using the ear alone is about 90%. (ii) we examined various feature extraction methods to learn the best features for head side view and auricle recognition including shape-based, namely Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF); and texture-based, namely Multi scale Local Binary Patterns (MLBP), Local Ternary Patterns (LTP). We determined that texture-based techniques perform better considering that the MLBP yielded the best performance with 90.20% rank-1 identification; and (iii) we evaluated the effect of different fusion schemes, at the image, feature, and score levels, on the recognition performance. Weighted Score fusion of face profile and ear has the best score with 91.14% rank-1 identification.

S. El-Naggar (✉) · T. Bourlai
West Virginia University, Morgantown, USA
e-mail: selnagga@mix.wvu.edu

T. Bourlai
e-mail: Thirimachos.Bourlai@mail.wvu.edu

A. Abaza
Cairo University, Cairo, Egypt
e-mail: aabaza@mix.wvu.edu

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_4

Introduction

Biometrics refers to the automatic measurement and analysis of individuals' distinctive physical and behavioral characteristics such as face, voice, iris and finger prints for authentication. Biometric systems are deployed in a wide variety of applications ranging from security applications, such as border control and national ID, to commercial applications, including network access and mobile devices. Face is one of the most popular biometric modalities [1]. This was motivated by the advantages of face recognition such as being non-intrusive, natural, passive and socially accepted [2].

Despite the great advances in face recognition technology, conventional face recognition systems depend mostly on the availability of full frontal face images in order to operate effectively. Most of the commercial face recognition systems typically detect pose variation as one of the preprocessing steps, and only when it is acceptable (frontal, or close to frontal) the system further process these images to establish human identity. Unfortunately, in real-life situations when identifying non-cooperative subjects in public spaces or unconstrained environments, like those encountered in surveillance applications, frontal face images may not be available. The only biometric that may be available for recognition is partial biometric information, like head side view.

Face profile, which is the side view of the face, provides biometric related informational content complimentary to the information provided by frontal view face images. Moreover, the outer ear or auricle, is part of the head side view and has demonstrated to have discriminative information across individuals; making it useful for recognition. According to [3], auricle doesn't seem to suffer from some of the problems inherent in the face (as an independent biometric modality), such as facial expression. Thus, ear-based recognition systems, under certain conditions, can effectively extend the capabilities of stand alone face recognition systems.

Assuming that profile face images are available (either directly or via restoration) and the image quality allows for ear or profile based authentication, here is a list of conditions/scenarios where head side view can be used for authentication:

1. Non-cooperative subjects in uncontrolled environments (surveillance systems), for example recognition of people entering rooms for home safety applications [4].
2. Drivers passing through security checkpoints [5].
3. Mobile users to unlock their phones [6].
4. Mug-shot, where datasets consist of one frontal face image and one side view face image per subject.

While both face profile and ear based recognition systems are important and have multiple applications, there is no clear definition for what a face profile is or which features in the head side view provide the most discriminant identity cues. The objectives of this work are to:

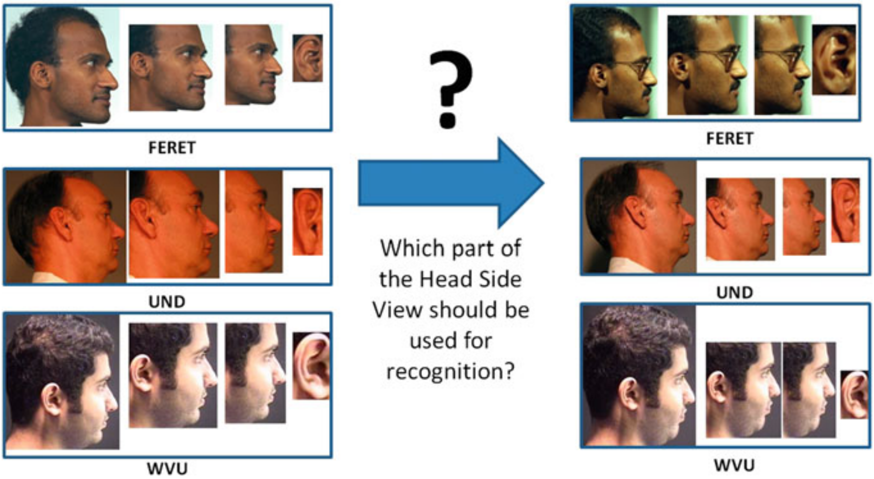

FERET
UND
WVU
?
Which part of the Head Side View should be used for recognition?
FERET
UND
WVU

Fig. 4.1 Samples of multiple parts of the head side view used for recognition performance evaluation which are: (i) full side view of the head, (ii) side view of the head without the hair part, (iii) side view of the head without the hair and the ear parts, and (iv) ear only. The gallery images are on the left while the probe images on are the right. Top: Sample from FERET dataset, middle: Sample from UND dataset, bottom: Sample from WVU dataset

- Examine which part(s) of the head side view is/are more beneficial for recognition: (i) full side view of the head (including hair), (ii) side view of the head without the hair region, (iii) side view of the head without the hair and the ear regions, or (iv) ear only (see Fig. 4.1).
- Compare the performance of various feature extraction techniques that are commonly used for face recognition, namely shape-based techniques such as Scale Invariant Feature Transform(SIFT) [7], Speeded Up Robust Features (SURF) [8]; and texture-based techniques such as Multi scale Local Binary Patterns (MLBP) [9], Local Ternary Patterns (LTP) [10].
- Determiner which of the various fusion scenarios of face profile and ear traits: at the image/sensor level, feature level, or score level yield the best performance results.
- Evaluate system performance (identification and verification) for all the aforementioned studies.

Performance evaluation studies are conducted using a set of popular head side and ear datasets, including the USTB dataset I [11], the UND dataset (collections E, and F) [12], the FERET [13, 14], and the WVU [15] datasets.

Please note that this work does not focus on proposing new matching algorithms using auricle or face profile images. It is an extended study on assessing the performance of auricle and head side view images under various scenarios in terms of the biometric components used, feature extraction methods and fusion approaches. To the best of the authors' knowledge, there is no analogue study published thus far that provides such analysis on the selected biometric modalities.

The rest of this chapter is organized as follows: section "Related Research" highlights some of the previous work on 2D ear recognition, face profile recognition and the fusion of face profile with ears to establish recognition. Section "Face Side/Ear Recognition Methodology" provides a brief overview of several feature extraction techniques that are used for ear and face recognition. Experimental results are presented in section "Experimental Results". Section "Case Study" gives an outline of our proposed approach for side view recognition. Finally, section "Conclusion and Future Work" presents conclusions and describes our future plans.

Related Research

In the literature, there is no clear definition for what is a face profile, or which features in the head side view provide the identity cues. This section discusses existing techniques for face profile recognition, ear recognition and an gives an overview of the existing research in fusion of face profile and ear for recognition.

Face Profile Recognition

Face profile recognition has been handled in the literature in two main approaches [16]: First, the probe image is a side view face image and the gallery image is a front view face image. In such case, the problem is a severe case of face recognition across pose, where the pose is 90°, Zhang and Gao [17] reviewed the problem with a survey of the techniques that had been proposed/used to handle it. Regardless, of the various attempts to overcome such a problem, it is still an unsolved challenge since the performance of face recognition systems degrade excessively with such pose angles. Second, the gallery and the probe images are side view face images. In this case the problem is considered a case of multi-view face recognition. Most of the face profile recognition methods utilize only the face profile contour line (silhouette) [18].

Silhouette based methods were first used for face profile recognition by Kaufman et al. [19] in 1967. Their work was followed by a lot of research for face profile recognition using silhouettes. Some determined fiducial points and used them to extract lines, angles, and areas as features [20–22]. Others used profile curve segments for matching [23, 24]. Ding et al. [25] used discrete wavelet transform to decompose the curve of the face silhouette. Then, they generated random forest models and used them for authentication.

Techniques that use only profile line have their advantages such as less complicity, memory usage, and maybe more robust to illumination changes. Unfortunately, they don't tolerate any pose variation and depend on clear images only. Additionally, they do not take advantage of the facial features or the texture information in the images [4].

Deep learning algorithms have been recently deployed in many artificial intelligence applications, including but not limited to, image classification, object detection/recognition and face recognition. Deep learning algorithms demonstrated distinguished performance which surpassed the state of the art for the above mentioned applications [26]. Krizhevsky et al. [27] developed a deep convolutional neural network for image classification that won the ImageNet challenge ILSVRC-2012.

Taigman et al. [28] developed a deep learning neural network for face recognition, they named it DeepFace. In a preprocessing step, they perform 3D alignment based on fiducial points. They used a large dataset of facebook images includes 4.4 million labeled faces from 4,030 people for training their neural network. Their experiments showed 97.35% accuracy using Labeled Faces in the Wild (LFW) dataset and 92.5% using the YouTube Faces (YTF) dataset. Schroff et al. [29] introduced FaceNet a ConvNet for face verification, recognition and clustering. To train their network they used a google dataset of 200 million face thumbnails for about 8 million identities. Their experiments showed 99.63% accuracy using (LFW) dataset and 95.12% using the (YTF) dataset. Parkhi et al. [30] assembled a dataset of 2.6 million face images for over 2.6K people of which approximately 95% are frontal and 5% profile face images. They used the ConvNet of [31] for face recognition and performed the triplet loss training. Their experiments showed 98.95% accuracy using (LFW) dataset and 97.40% using the (YTF) dataset.

In a recent application of convolutional neural networks, Zhang and Mu [32] proposed a technique involving Multiple Scale Faster Region-based Convolutional Neural Network for ear detection. They used information related to ear location context to locate ear accurately and eliminate the false positives. They tested their system on three different datasets: web ear, UBEAR and UND-J2 and their experiments showed 98.01, 97.61, and 100% accuracy respectively. They also examined the system with a test set of 200 web images (with variable photographic conditions), and achieved a detection rate of approximately 98%.

Ear Recognition

For ear recognition, Abaza et al. provided an overview of the existing ear recognition techniques [3]. The popular Principal Component Analysis (PCA) representation was first used for ear recognition by Chang et al. [33] who introduced the concept of Eigen-Ear. Following, PCA had been widely used in the literature as a base reference.

Dewi and Yahagi [34] used Scale-Invariant Feature Transform (SIFT) [35] feature descriptor for ear recognition. In their work, they classified the owner of an ear by calculating the number of matched key points and their average square distance. While Kisku et al. [36] used SIFT for colored ear images. They segmented the ears in decomposed color slice regions of ear images, then they extracted SIFT key-points and fused them from all color slice regions.

Local Binary Patterns (LBP) were combined with wavelet transform for ear recognition in [37]. Wang et al. [38] decomposed ear images by a Haar wavelet transform and then applied Uniform LBP simultaneously with block-based and multi-resolution methods to describe the texture features.

Face Profile and Ear

There has been a low amount of attention in the literature for face profile and ear fusion for authentication. Gentile et al. [5] introduced a multi-biometric detection system that detects ear and profile candidates independently. For all profile candidates if an ear exists, that is contained inside the profile, that profile region is labeled as a true profile.

While for identification, Yuan et al. [39] used face profile images that includes the ear (assuming fusion at the sensor/image level). They applied Full Space Linear Discriminant analysis (FSLDA). Xu and Mu [40] used the same technique FSLDA for face profile as a uni-modal and the ear as another uni-modal. They carried out decision fusion using combination methods of Product, Sum, and Median rules according to the Bayesian theory and a modified vote rule for two classifiers. Later, Pan et al. [41] modified the FSLDA technique by using kernels of the feature vectors. They fused the face profile and ear at the feature level and applied the Fisher Discriminant Analysis (FDA). Face profile and ear were used in a PCA-based recognition system for robotic vision [42]. In the system, if either the ear or the face of a person was recognized successfully, it is considered a correct identification of the subject in a decision level fusion.

In a close work, Rathorea et al. [43] proposed fusing ear information with profile face information to enhance recognition performance. They used SURF for feature extraction. In their method, for each of the face profile and the ear they extracted three sets of SURF features. Each set was obtained after applying one of three image enhancement techniques. They reported their results on three data bases individually UND-E, UND-J2 and IITK.

Face Side/Ear Recognition Methodology

The goal of this work is to investigate challenging biometric scenarios when only images of face side view are available for recognition rather than frontal face images captured under controlled conditions. Our objective was to assess the following, (a) Which part of the head side provides the identity cues? (b) Which part of the side view image contains the discriminative information necessary for identification or verification different scenarios?

We developed an automated system for ear detection using Haar features arranged in a cascaded Adaboost classifier [44]. The system is fast and robust for partial occlusion and rotation. We tested the system using 2113 profile images for 448 different cases and achieved 95% correct ear detection. In this study, and to avoid error carried from the detection stage, we manually cropped multiple parts from the head side view image which are:

1. Full side view of the head.
2. Side view of the head without the hair part.
3. Side view of the head without the hair and the ear parts.
4. Ear only.

In our study we are also experimenting, which feature extraction method will provide the most distinctive representation for each of the fragments of the head side view?

The face side view provides salient and texture rich information in counter to the external ear that has morphological components. Such structure motivated us to examine both, shape-based and texture-based, standard feature descriptors that are commonly used for frontal face images.

1. *Shape-based* examined techniques are:
   - Scale invariant feature transform (SIFT).
   - Speeded-Up Robust Features (SURF).
2. *Texture-based* description techniques:
   - Multilevel Local Binary Patterns (MLBP).
   - Local Ternary Patterns (LTP).

The performance of the aforementioned techniques was evaluated for identification and recognition scenarios. Moreover, we evaluated biometric fusion methods at various levels. First, face profile/ear fusion at the image or sensor level in the full side view images and the face side view images without the hair part. Second, fusion at the feature level using simple concatenation rule of face profile-ear features. Third, fusion at the score level which was accomplished by consolidating scores for face profile-ear using both normal average rule and weight sum rule. An over-view of the evaluation studies performed with different combinations is shown in Fig. 4.2.

SIFT Description

The scale invariant feature transform (SIFT) is a shape-based algorithm for extracting highly distinctive invariant features. Lowe's SIFT algorithm [35], consists of four major stages:

- *Scale-space extreme detection*: where a difference-of-Gaussian function is applied to the image to identify candidate points that are invariant to scale and orientation, as follows:

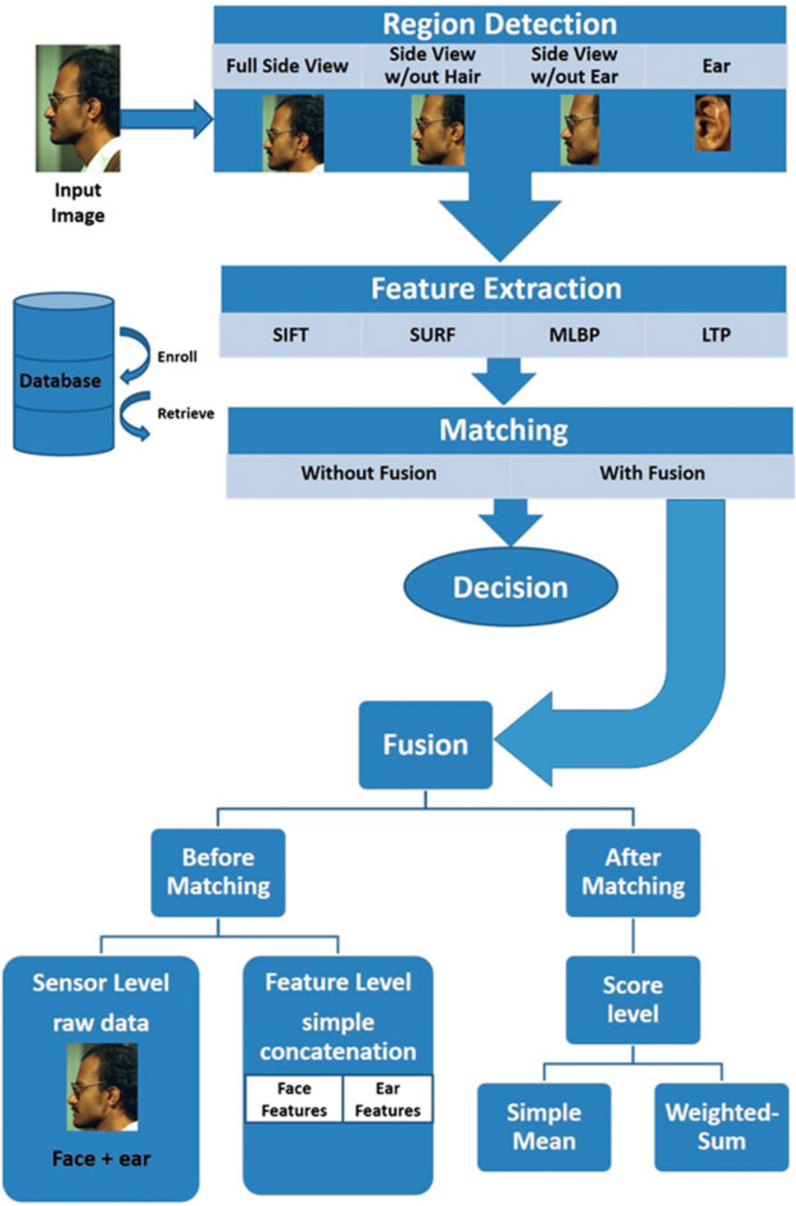


Fig. 4.2 Overview of the different scenarios experientially evaluated. The recognition performance was evaluated for multiple parts of the head side view: (i) full side view of the head, (ii) side view of the head without the hair part, (iii) side view of the head without the hair and the ear parts, and (iv) ear only. Multiple feature extraction techniques were tested: (i) SIFT, (ii) SURF, (iii) MLBP. (iv) LTP. Fusion applied at: (i) sensor/image level, (ii) feature level, and (iii) score level

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y), \tag{4.1}$$

where $I(x, y)$ is the original image, and $k\sigma$, σ refer to different Gaussian-blur separated by a constant multiplicative factor k.

- *Key point localization*: it rejects points having low contrast (sensitive to noise) or are poorly localized along an edge. The initial implementation of SIFT approach [45] located key-points at the location and scale of the central sample point. Lowe [35] used the Taylor expansion (up to the quadratic terms):

$$D(x) = D + \frac{\partial D^T}{\partial x} x + 0.5x^T \frac{\partial^2 D}{\partial x^2} x, \tag{4.2}$$

where D and its derivatives are evaluated at the sample point and $x = (x, y, \sigma)^T$ is the offset from this point.

For stability, edge responses are eliminated. A poorly defined peak in the difference-of-Gaussian function will have a large principal curvature across the edge but a small one in the perpendicular direction. The principal curvatures can be computed from a 2×2 Hessian matrix H, computed at the location and scale of the key-point:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}, \tag{4.3}$$

The eigenvalues of H are proportional to the principal curvatures of D.

- *Orientation assignment*: where a gradient orientation histogram is computed in the neighborhood of each key-point, where histogram peaks correspond to dominant orientations.
- *Key point descriptor*: for each selected key-point orientation, a feature descriptor is computed as a set of orientation histograms.

Dense SIFT [46] extracts local feature descriptors at regular image grid points yielding a dense description of the face images, while normal sift extract feature descriptions at the locations determined by Lowe's algorithm [45]. Dense SIFT yielded inferior performance compared to SIFT; hence we decided to proceed with the original SIFT technique.

SURF Description

Speeded-Up Robust Features (SURF) is based on similar properties to SIFT. Bay et al. [47] used basic Hessian-matrix approximation for interest point detection. They detected blob-like structures at locations where the Hessian-matrix determinant is maximum. Given a point $x = (x, y)$ in an image I, the Hessian matrix is defined as:

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix}, \tag{4.4}$$

where $L_{xx}(x,\sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2}$ with the image I in point x, and similarly for $L_{xy}(x,\sigma)$ and $L_{yy}(x,\sigma)$.

SURF descriptor [47], describes the distribution of the intensity content within the interest point neighborhood as follows:

- Orientation Assignment: To be invariant to image rotation, a reproducible orientation for the interest points is identified.
- Descriptor based on Sum of Haar Wavelet Responses: The region is split up regularly into smaller 4×4 square sub-regions. For each sub-region, d_x the Haar wavelet response in horizontal direction and d_y the Haar wavelet response in vertical direction respectively is calculated. Then, the wavelet responses d_x and d_y are summed up over each sub-region. To bring in information about the polarity of the intensity changes, the sum of the absolute values of the responses, $|d_x|$ and $|d_y|$ are also extracted.

MLBP Based Description

Local Binary Patterns (LBP) operator is a texture descriptor that quantify the intensity patterns in local pixel neighborhood patches, and have been used for face recognition in [48]. They have shown the LBP operator to be highly discriminative and computationally efficient. Using LBP operator for ear recognition is based on the description of ears as a composition of micro-patterns. The basic LBP operator assigns a decimal value to each pixel in the image by thresholding ($P = 8$) neighbor pixels at distance ($R = 1$), as follows:

- For a given input image pixel I_c and its 8 neighbors I_p,
- Each neighbor pixel greater than or equal to the center pixel is assigned 1 otherwise it is assigned 0:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(I_p - I_c)2^P, \tag{4.5}$$

where $s(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{if } \mathit{otherwise}. \end{cases}$

- These binary values are arranged to form a binary number (01110010), which is transferred to a decimal equivalent (114).
- The histogram (H) of these decimal values represents the feature vector.

Ojala et al. [9] defined a local binary pattern uniform if the binary number contains at most two bitwise transitions from 0 to 1 or vice versa. For example: 00000000, and 11000011 are considered uniform. This feature selection method reduced the number of features, in case of 8-bin histogram, from 2^8 to 59. Multi-Scale Local Binary Patterns (MLBP) concatenated the histogram computed by LBP descriptors at four different radii $R = 1, 3, 5, 7$, while keeping $P = 8$, to yield better performance. Block-based Local Binary Patterns divided the face or ear image into a set of blocks that can be overlapped.[1]

To measure the similarity between the probe histogram H^p and gallery histogram H^g generated by LBP operator, the Chi-square distance was used:

$$S_{Chi}(H^p, H^g) = \Sigma_{j,i}\omega_j * \frac{(H^p_{i,j} - H^g_{i,j})^2}{(H^p_{i,j} + H^g_{i,j})} \tag{4.6}$$

where i and j refer to the ith bin in histogram corresponding to the jth block, and ω_j is the weight for block j.

LTP Based Ear Description

Local Ternary Patterns (LTP) operators extends LBP to 3-valued codes [10], in which gray levels, in a zone of width $\pm t$ around a centered value, are quantized to zero; ones above this are quantized to +1 and ones below it to −1, i.e. the indicator s(u) is replaced by a 3-valued, as follows:

- For a given input image pixel I_c and its 8 neighbors I_p,
- Each neighbor pixel greater than the center pixel I_c plus t is assigned, or less than the center pixel minus t is assigned to −1, otherwise it is assigned 0:

$$LTP_{P,R} = \sum_{p=0}^{P-1} s(I_p - I_c)2^P, \tag{4.7}$$

where $s(x) = \begin{cases} 1 & \text{if } x > t \\ 0 & \text{if } -t \le x \le t \\ -1 & \text{if } x < -t \end{cases}$

- These ternary values are arranged to form a ternary number (01"−1" "−1"1110). This ternary number is transferred into two binary numbers (01001110, 00110000), which are then transferred into two decimal equivalents (92, 48).
- Two separate channels of LBP descriptors, for which separate histograms of these decimal values, forms the feature vector.[2]

[1] Experimentally overlapped MLBP, 24 × 24 pixels patches that overlap by 12 pixels, was proven to yield the best performance.

[2] Experimentally overlapped LTP, 24 × 24 pixels patches that overlap by 12 pixels, was proven to yield the best performance.

To measure the similarity between the probe histogram H^p and gallery histogram H^g generated by LTP operator, the Chi-square distance was also used.

Experimental Results

This section starts with a description of various face side and ear datasets that we used in our experiments, followed by an explanation of the experiments performed, the results obtained, and a discussion of the results. We designed the experiments to examine:

1. Which part of the face side should be used for recognition and whether the ear should be included?
2. Which feature extraction method is more effective for face side/ear recognition?
3. Which fusion scenario for ear and side face biometrics can improve the identification performance?

Ear Datasets

We composed a heterogeneous test dataset that consists of images from three different datasets, UND, FERET, and WVU. This was to overcome the limited size of the available datasets. Additionally, we used a fourth dataset, the USTB dataset, for parameter estimation of those feature extraction methods that required training phase. Table 4.1 shows the components of the test/training dataset that we used:

1. The University of Notre Dame (UND) dataset[3] The UND dataset consists of multiple collections for face and ear modalities.
   - Collection E contains 464 left face profile (ear) images of 114 subjects. From this collection, we used the images of 102 subjects to maintain 2 images per subject.
   - Collection F contains 907 right face profile (ear) images of 286 subjects. From this collection, we used images of 285 subjects to maintain 2 images per subject.
2. FERET dataset [13, 14]: The FERET dataset was part of the Face Recognition Technology Evaluation (FERET) program. The dataset was collected in 15 sessions between August 1993 and July 1996. It contains 1564 sets of images for a total of 14126 images that includes 1199 individuals and 365 duplicate sets of images. For some individuals, images were collected at right and left profile (labeled pr and pl). From this dataset, we used left face profile (ear) images of 115

[3] http://www3.nd.edu/~cvrl/CVRL/Data_Sets.html.

Table 4.1 Datasets used in our experiments

Data set	Left face side	Right face side
UND, collection E	–	102
UND, collection F	285	–
FERET	115	125
WVU	60	58
Test set	460	285
Training USTB	–	60

subjects, and right face profile (ear) images of 125 subjects to maintain 2 images per subject.

3. WVU dataset [15]: The WVU ear dataset consists of 460 video sequences for about 400 different subjects and multi-sequences for 60 subjects. Each video begins at the left profile of a subject (0°) and terminates at the right profile (180°) in about 2 min. This dataset has 55 subjects with eyeglasses, 42 subjects with earrings, 38 subjects with partially occluded ears, and 2 fully occluded ears. We used 60 left face profile (ear) images of 60 human subjects, and 58 right face profile (ear) images of 58 human subjects.
4. The University of Science and Technology Beijing (USTB) datasets[4]: The USTB Dataset consists of several ear image datasets. Image Dataset I contains 180 images of 60 subjects. The ear images in the USTB dataset I are vertically aligned. We used this dataset for estimating the parameters of the feature extraction techniques; we call it training set. For example for the MLBP, the USTB was used for the estimation of: the size of the local windows, the overlap between the local windows and the number of sample points.

Performance of Various Feature Extraction Methods

Using the test dataset, we evaluated the following feature extraction methods for the various parts of the face side view:

- Scale Invariant Feature Transform (SIFT).
- Speeded Up Robust Features (SURF).
- Multi-scale Local Binary Patterns (MLBP).
- Local Ternary Patterns (LTP).

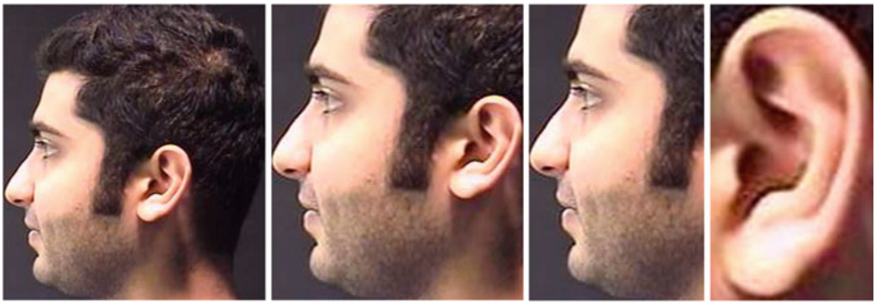
Biometric systems typically operate in either identification mode or verification mode. In identification, it determines the identity from a database while in verification it confirms the identity claimed. The performance of a biometric matcher in

[4] http://www1.ustb.edu.cn/resb/en/index.htm.

identification mode is based on the Cumulative Match Characteristic (CMC) curve [49]. The CMC curve depicts the probability of obtaining the correct identity in the top n ranks cumulatively. Differently, the performance of a biometric matcher in verification mode is based on the Receiver Operating Characteristic (ROC) curve. The ROC curve depicts the percentage of False Accept Rate (FAR) versus the percentage of False Reject Rate (FRR) at varying threshold. We utilized the matching scores of the different feature extractors to generate the CMC and ROC curves for performance comparison, as well as assessment of which part in the face side view is more useful for personal authentication. Figure 4.3 shows the different representations of the face side view including (full side view of the face, side view without hair, and side view without ear) as well as ear only. Table 4.2 shows rank-1 of the identification experiment. Figures 4.4, 4.6, 4.8 and 4.10 show the CMC curves of the SIFT, SURF, MLBP and LTP feature extraction techniques, respectively, for various parts of the face side view. Figures 4.5, 4.7, 4.9 and 4.11 show the ROC of SIFT, SURF, MLBP and LTP feature extraction techniques, respectively.

The findings of these experiments are as follows:

- We attribute the performance difference between the right and the left side views to the variance in the size of the dataset and the variance in the lighting conditions between the gallery and probes for the UND dataset.
- Face side view images that include the ears, which can be considered as fusion of face profile and ear at the image level, provides better identification accuracy than that of the face profile images without the ears; hence it is recommended to keep the ear region when using the face side view for recognition. We lay the enhancement in performance to the morphological components of the ear, which make the ear shape information retrievable from the shadow information.
- Side view face, including the ear and without the hair part, outperforms the complete head side view including the hair region; hence it is recommended to crop out the hair region when using the face side view for recognition. We attribute the decrease in the performance to the noise provided by the hair region; plus it is easy to change the hair cut/style, which may mislead the system.



(a) Complete *(b) WEar* *(c) W/OEar* *(d) Ear*

Fig. 4.3 Examples of head side part and the ear

Table 4.2 Comparison of identification (rank-1) rate several techniques

Left set (R1%)	SIFT	SURF	MLBP (O)	LTP (O)
Side	71.30	69.13	52.39	51.30
Profile (W ear)	72.61	75.65	65.00	59.13
Profile (W/O ear)	61.30	56.74	60.87	54.78
Ear	33.48	44.35	**88.48**	81.30
Right set (R1%)	SIFT	SURF	MLBP (O)	LTP (O)
Side	87.37	87.37	80.70	80.70
Profile (W ear)	85.61	87.72	82.81	81.40
Profile (W/O ear)	76.14	73.68	81.40	80.00
Ear	63.86	65.96	**92.98**	91.93
Average (R1%)	SIFT	SURF	MLBP (O)	LTP (O)
Side	77.45	76.11	63.09	62.55
Profile (W ear)	77.58	80.27	71.81	67.65
Profile (W/O ear)	66.98	63.22	68.46	64.43
Ear	45.10	52.6	**90.20**	85.37

- The performance of the ear region alone using MLBP and LTP provides better identification performance compared to side view including the ear. We attribute the low performance of the SIFT and SURF techniques to failure in enrollment (in other words ear images has no or insufficient extracted SIFT/SURF points), which explains the cut off in the ROC curves in Figs. 4.5, 4.7, 4.9 and 4.11.

Performance of Face Profile and Ear Fusion

Fusion is combining information from multiple biometric modalities or biometric systems. The integration of the information from multiple sources is more likely to provide better performance to the biometric system, which means more stable systems that match real-world applications. Fusion can be applied at the sensor/image, feature, match score, and decision levels, as well as rank level in case of identification mode [50]. In this experiment, we consider the following fusion levels:

- Side view image including ear can be considered as fusion of face profile and ear at the image/sensor level.
- Fusion at the feature-level using simple concatenation rule, where we consider the same features from face profile and ear; i.e. MLBP for both modalities. Table 4.3 shows rank-1 performance for fusion at the feature level.
- Fusion at the score-level, match scores output by face profile and ear matchers are consolidated. This approach has been widely used since match scores are easy to access and combine. However, match scores output by different biometric matchers

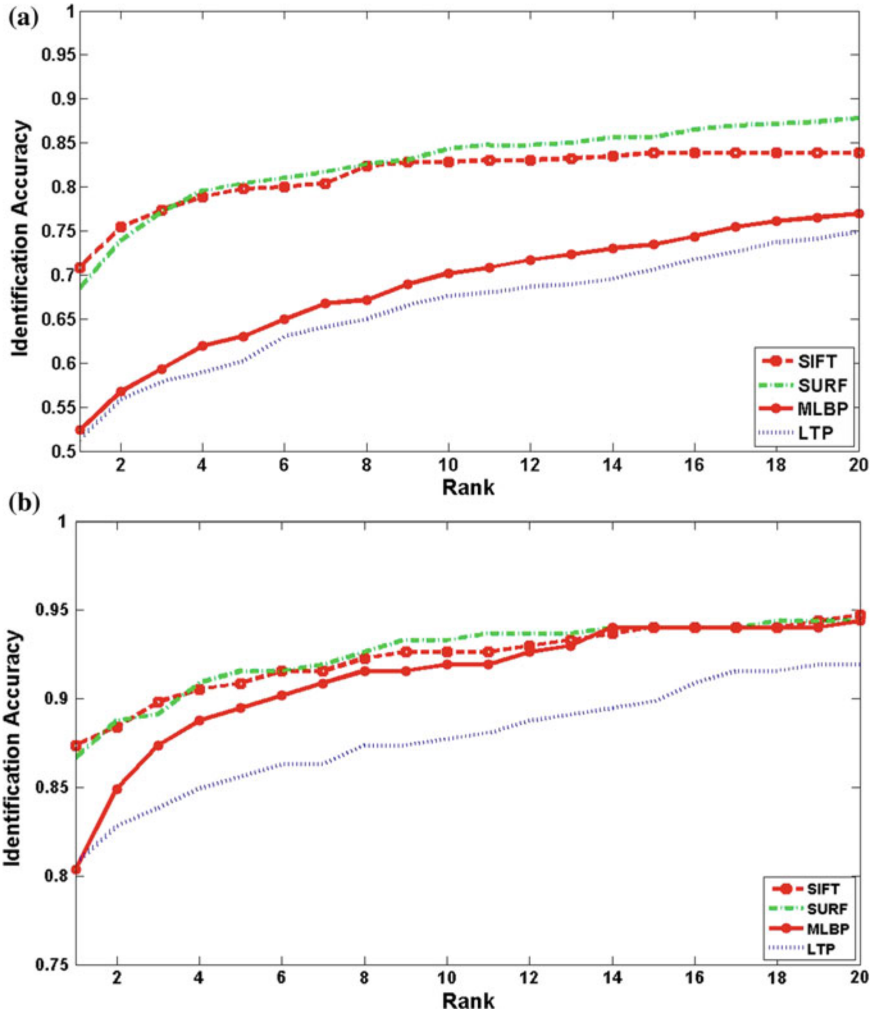


Fig. 4.4 Cumulative Match Characteristic (CMC) curve of several techniques (SIFT, SURF, MLBP, and LTP) for **a** left side images and **b** right side images. The horizontal axis of the CMC represents rank n, and the vertical axis represents the probability of obtaining the correct identity in the top n positions cumulatively

need a normalization step. Several integration rules can be used to implement score level fusion. A fusion rule which is commonly used in the literature is the simple mean formulated by the following formula $S_{mean} = (\sum_{k=1} s_k)/K$, where s_k is the match score output by the kth matcher. Another integration rule is the Weighted-Sum (WS) rule, where equal weights corresponds to a simple mean. To tune the weights employed in the weighted-sum rule, an experiment is carried out to find the amount of contribution of face profile and ear in the identification

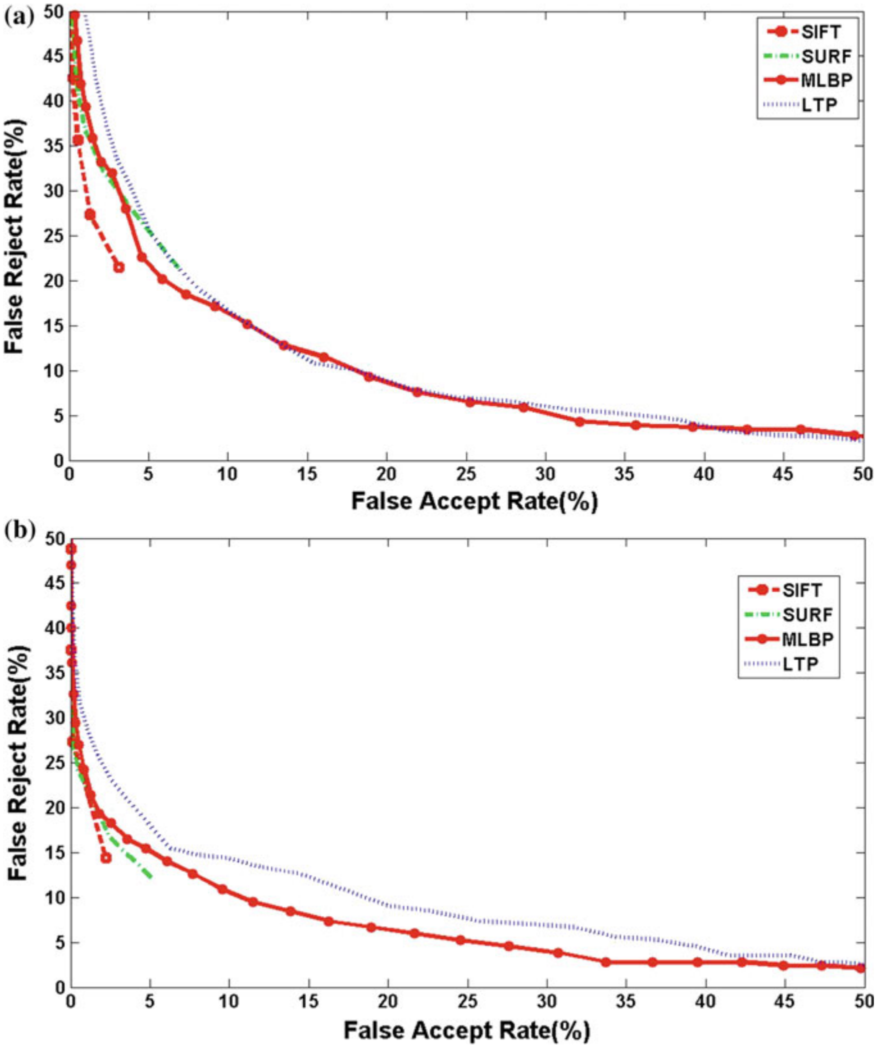


Fig. 4.5 Receiver Operating Characteristic (ROC) curve several techniques (SIFT, SURF, MLBP, and LTP) for **a** left side images and **b** right side images. The horizontal axis of the ROC represents False Accept Rate (%), and the vertical axis represents the False Reject Rate (%)

procedure $S_{WS} = \alpha S_{fp} + \beta S_{ear}$, where α is the face profile weight and β is the ear weight and $\alpha + \beta = 1$.

The findings of these experiments are as follows:

- Fusion at the score level, which is the most common approach in multibiometric systems [50], proved to yield better performance compared to fusion at the sensor/image and feature levels. We attribute the worse performance at the

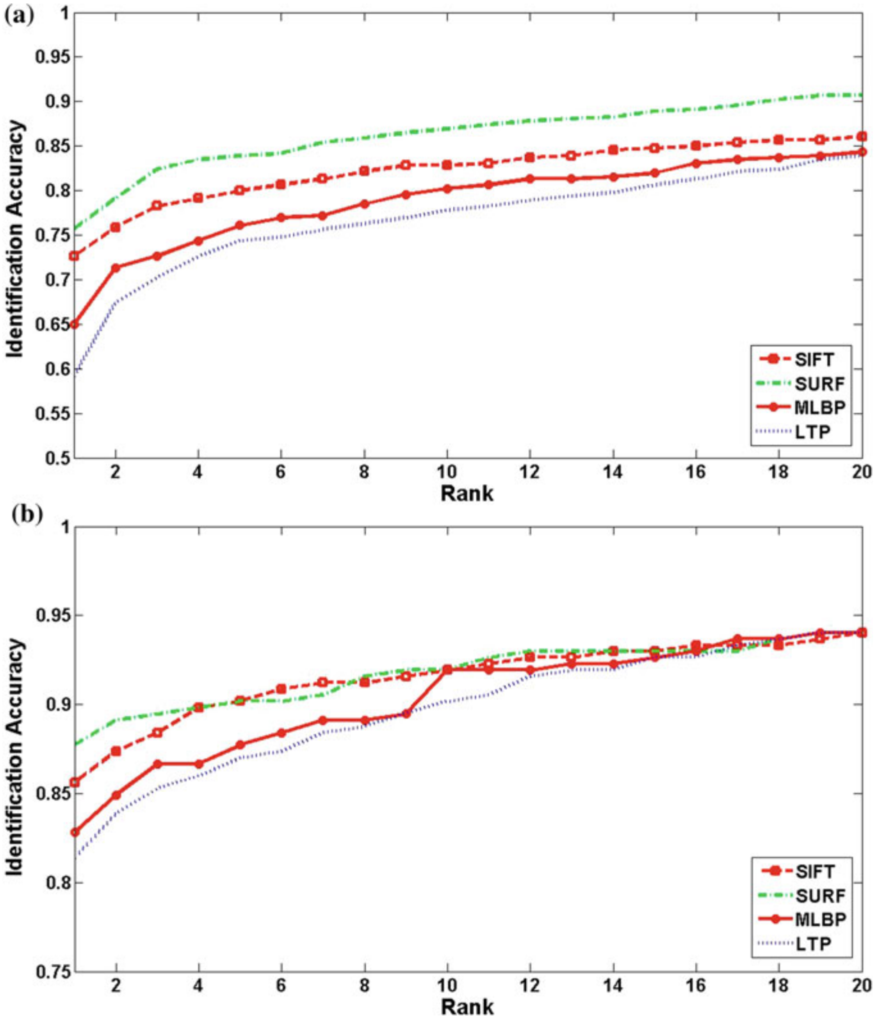


Fig. 4.6 CMC curves for **a** left profile images with ear and **b** right profile images with ear

feature level to the inferior performance of the selected fusion rule (simple concatenation).

- Table 4.4 shows that (i) for the texture-based techniques (MLBP and LTP), assigning more weight to the ear score enhances the overall system performance; (ii) for the shape-based techniques (SIFT and SURF), assigning more weight to the face profile score enhances the overall system performance; and (iii) suitable fusion of ear and face profile has synergy (i.e. it yielded an overall performance better than the simple addition of the two modalities).

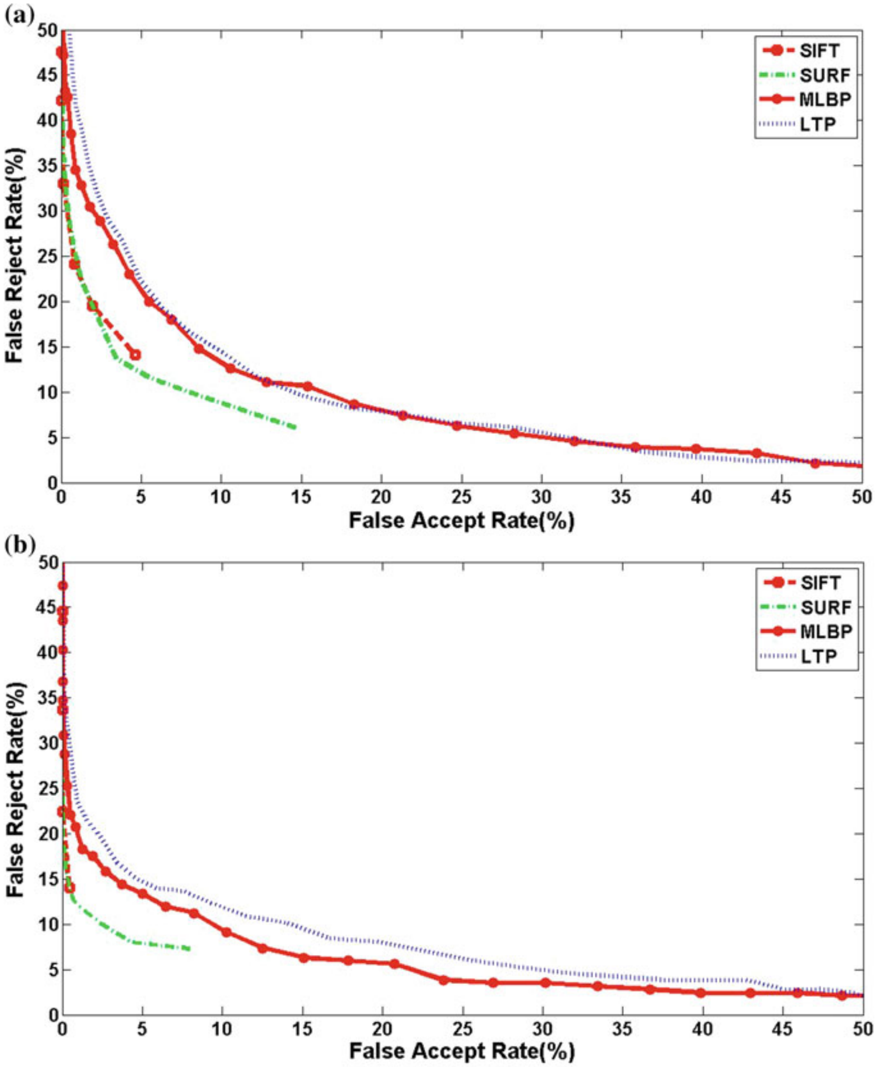


Fig. 4.7 ROC curves for **a** left profile images with ear and **b** right profile images with ear

Case Study

The objective of this section is to present an outline of the suggested approach for biometric authentication in the scenarios when only face side view is available, based on the analysis of the previous performance evaluation experiments. The suggested approach is as follows:

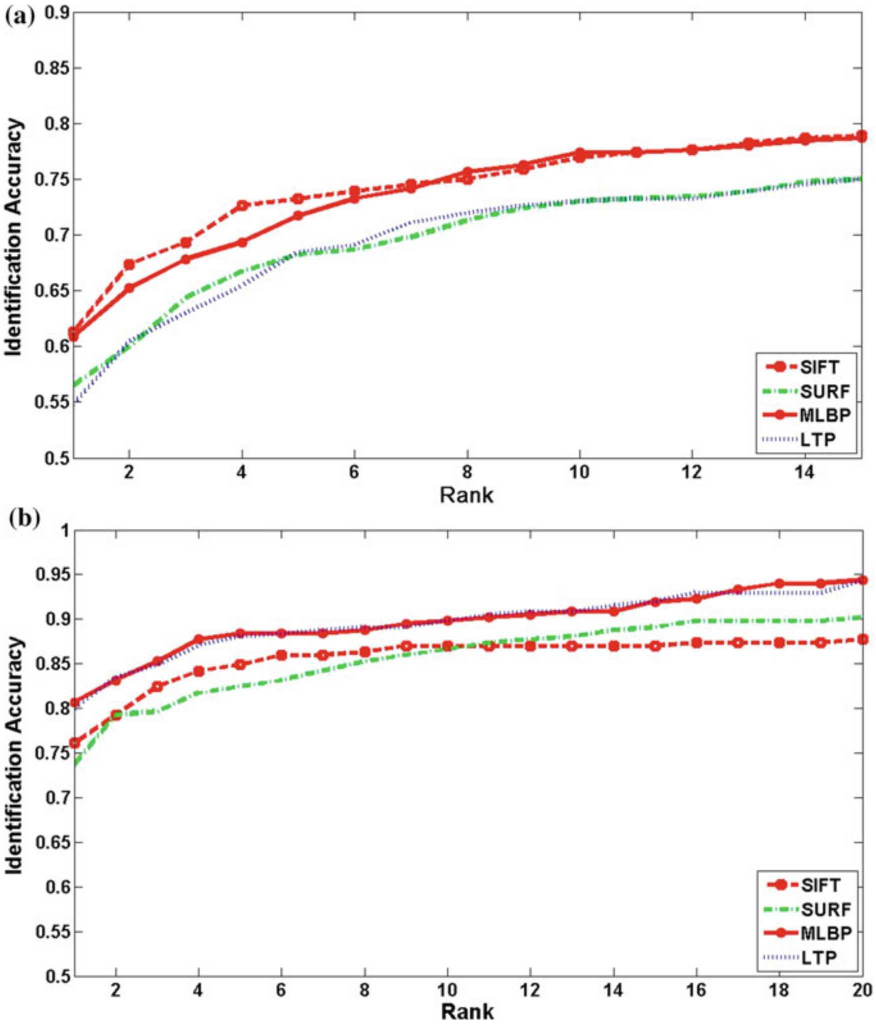


Fig. 4.8 CMC curves for **a** left profile images without ear and **b** right profile images without ear

1. Detect the ear region and the face profile without the hair/ear region, then crop them from the original image.
2. Use MLBP for each component feature's representation. Calculate the distance between the input pattern and the patterns enrolled in the database to generate the score matrix for each component individually.
3. Perform Weighted-Sum fusion with giving high weight to the ear scores ($W = 0.75$) and low weight to the face profile score ($W = 0.25$). Use the final output determine the identity of the candidate.

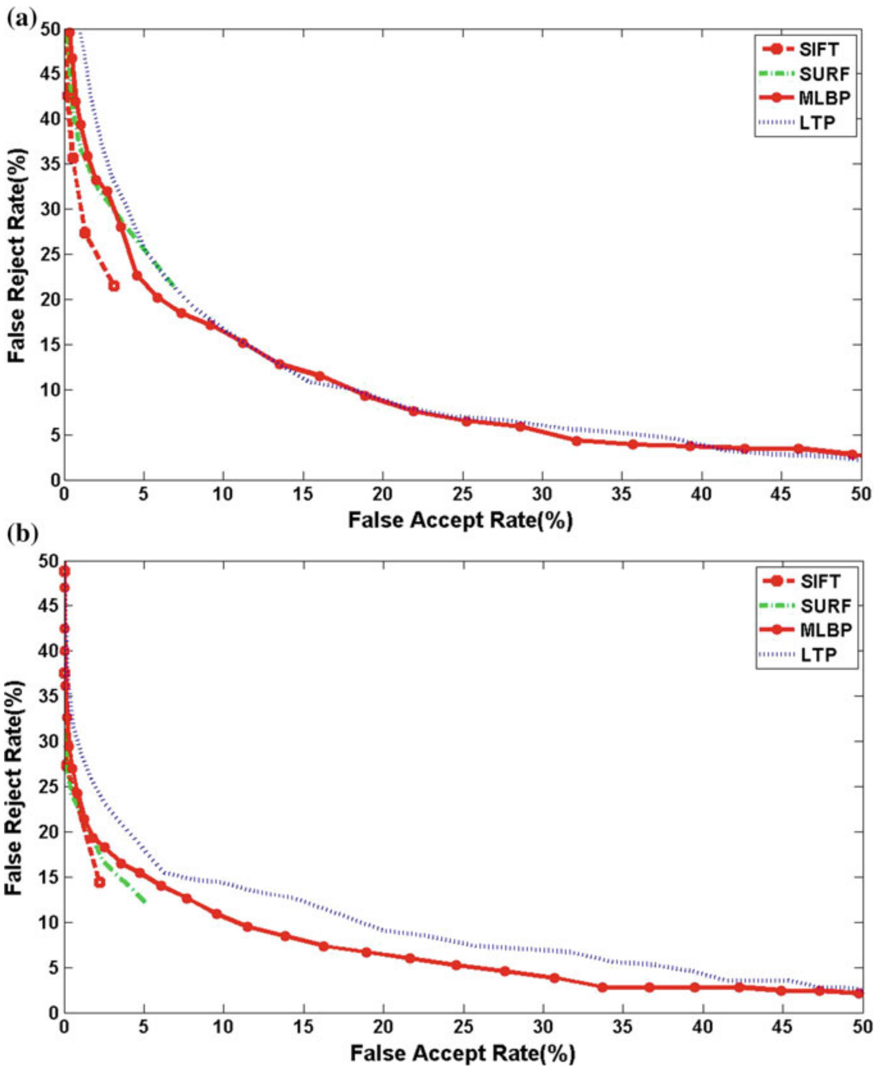


Fig. 4.9 ROC curves for **a** left profile images without ear and **b** right profile images without ear

Figure 4.12 shows an overview of the proposed side-view face recognition approach. Figure 4.13 shows the CMC curves for the proposed approach on our dataset and Table 4.4 shows rank-1 results.

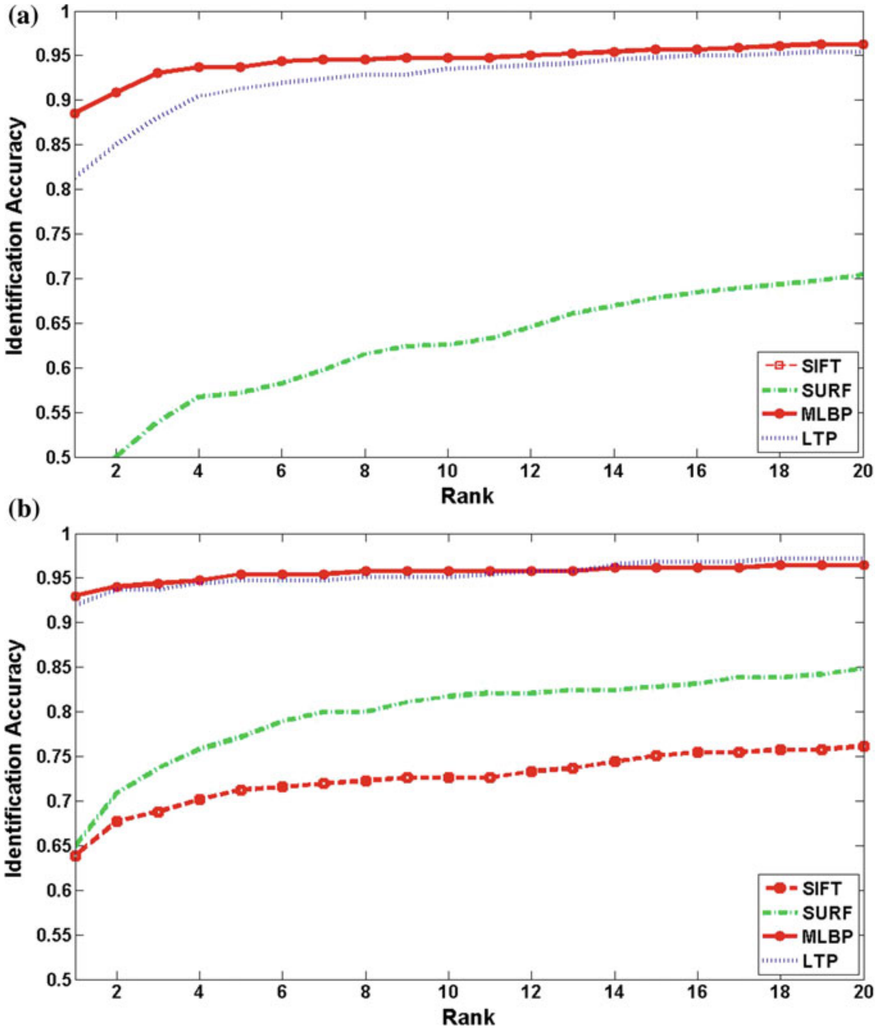


Fig. 4.10 CMC curves for **a** left ear images and **b** right ear images

Conclusion and Future Work

In this paper, we presented a study on the face side view and ear recognition by comparing the performance of several feature extraction methods using different parts of the face side view. This work does not focus on proposing new matching algorithms, which is an area for future work, but provides performance evaluation for multiple scenarios, components, approaches and databases. This study includes examining different parts of the face side view versus using the ear alone, comparing the performance of various feature extraction techniques that are commonly used for face

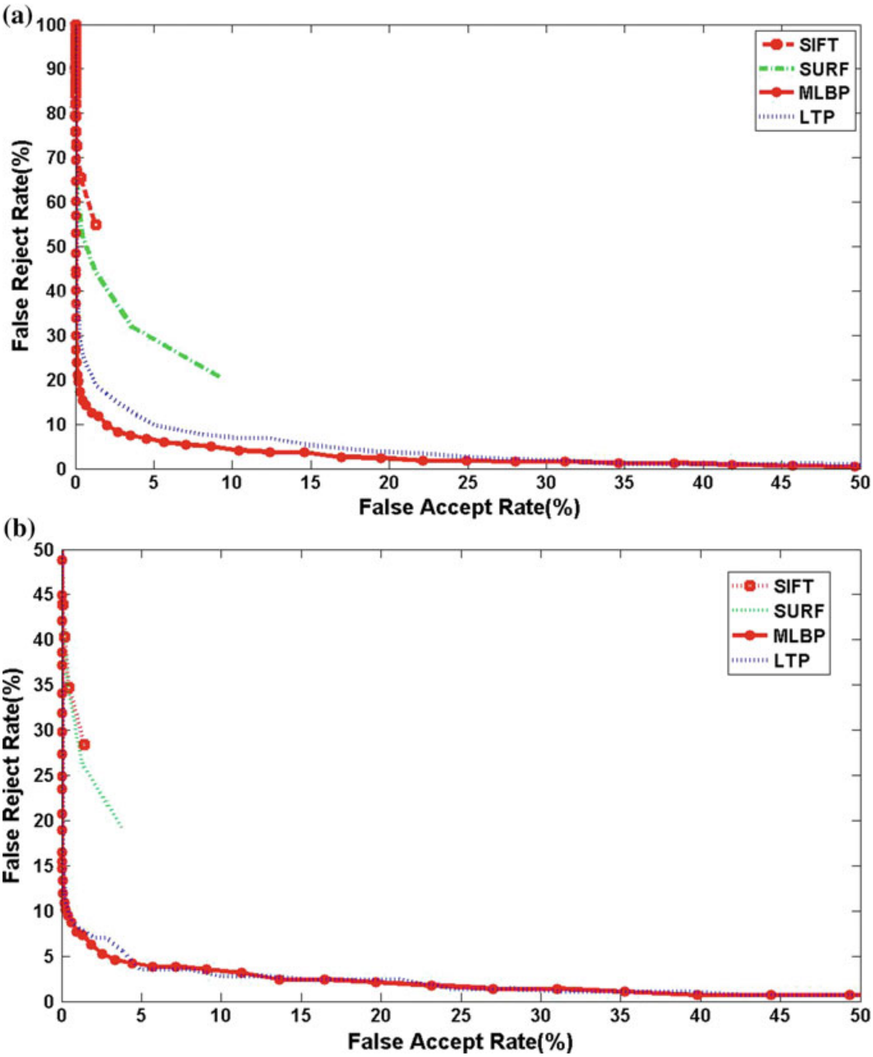


Fig. 4.11 ROC curves for **a** left ear images and **b** right ear images

recognition, and studying various scenarios for fusion of the face profile and ear traits at the sensor/image, feature, and score levels.

The main findings of the paper are the following:

- Ear provides main features in the side view, in terms of identity cues, where the rank-1 identification performance using the ear only reached about 90%.
- Texture-based techniques (MLBP 90.20%, LTP 85.37%) yielded better performance using the ear region only in comparison to side profile, while shape-based techniques (SIFT 77.58%, SURF 80.27%) yielded better performance using face

Table 4.3 Fusion at the feature level

Left set (R1%)	SIFT	SURF	MLBP (O)	LTP (O)
Profile (W/O) Ear	61.30	56.74	60.87	54.78
Ear	33.48	44.35	88.48	81.30
Fusion	29.35	**70.65**	64.13	59.57
Right set (R1%)	SIFT	SURF	MLBP (O)	LTP (O)
Profile (W/O) Ear	76.14	73.68	80.70	80.00
Ear	63.86	65.96	92.98	91.93
Fusion	60.35	**83.16**	82.11	81.75
Average (R1%)	SIFT	SURF	MLBP (O)	LTP (O)
Profile (W/O) Ear	66.98	63.22	68.46	64.43
Ear	45.10	52.62	90.20	85.37
Fusion	41.21	**75.44**	71.01	68.06

Table 4.4 Fusion at the score level

Left set (R1%)	SIFT	SURF	MLBP (O)	LTP (O)
Profile (W/O) Ear	61.30	56.74	60.87	54.78
Ear	33.48	44.35	88.48	81.30
Simple mean	56.74	65.43	86.52	77.83
WS ($\alpha = 0.25$, $\beta = 0.75$)	50.22	55.00	**89.78**	83.48
WS ($\alpha = 0.75$, $\beta = 0.25$)	66.30	69.13	78.04	66.52
Right set (R1%)	SIFT	SURF	MLBP (O)	LTP (O)
Profile (W/O) Ear	76.14	73.68	80.70	80.00
Ear	63.86	65.96	92.98	91.93
Simple mean	77.89	78.95	92.28	90.18
WS ($\alpha = 0.25$, $\beta = 0.75$)	76.14	73.68	**93.33**	92.63
WS ($\alpha = 0.75$, $\beta = 0.25$)	81.75	83.51	89.82	86.32
Average (R1%)	SIFT	SURF	MLBP (O)	LTP (O)
Profile (W/O) Ear	66.98	63.22	68.46	64.43
Ear	45.10	52.62	90.20	85.37
Simple mean	64.83	70.60	88.72	82.55
WS ($\alpha = 0.25$, $\beta = 0.75$)	60.14	62.15	**91.14**	86.98
WS ($\alpha = 0.75$, $\beta = 0.25$)	72.21	74.63	82.55	74.09

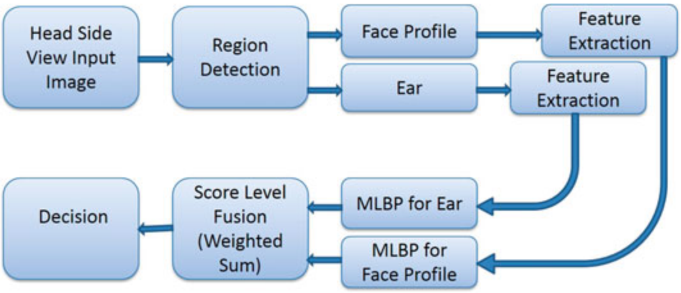


Fig. 4.12 An overview of the proposed side-view face recognition system

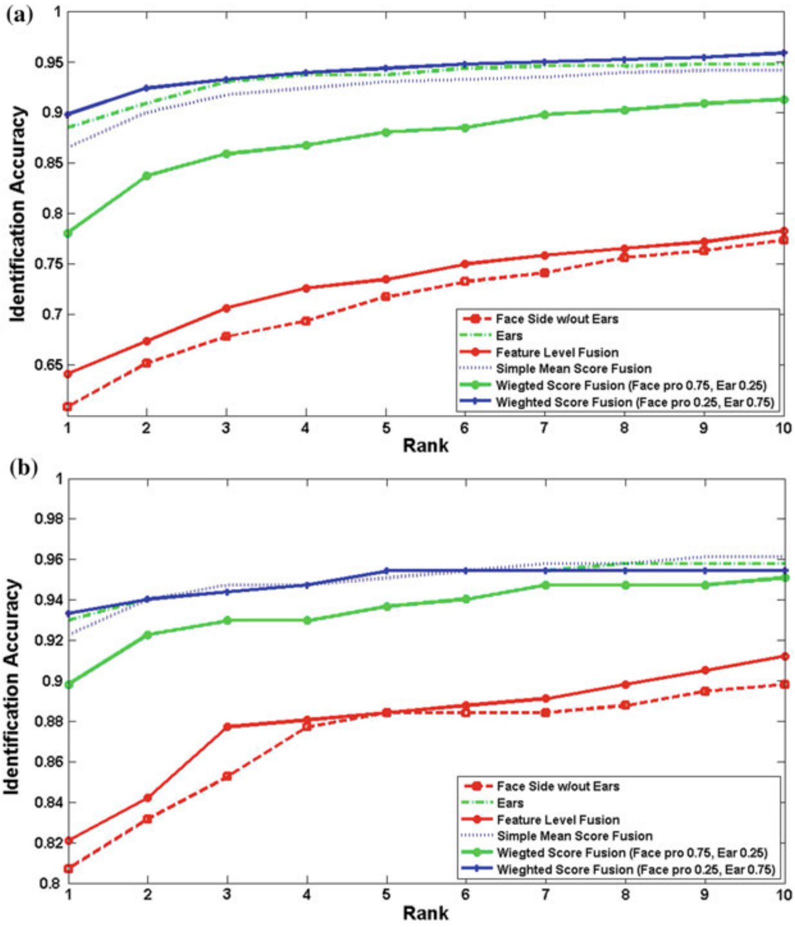


Fig. 4.13 CMC curves for **a** left ear images and **b** right ear for different fusion scenarios

profile with ear in comparison to using ear region alone. Overall MLBP using the ear region only yielded the best performance.

- Suitable fusion of side profile and ear has a synergic power (i.e. it yielded an overall performance better than simple addition of the two modalities). Overall multibiometric system of face profile and ear yielded 91.14% rank-1 identification on our heterogeneous test dataset that consists of 460 left side view and 285 right side view.

Future work is to assess other points regarding the performance of the face side: (i) aligning the face profile using the ear and nose locations, in a similar way to aligning the frontal face where the eye location is commonly used; (ii) testing the required resolution for the side view, and how this resolution is correlated to the extracted features; (iii) exploring new feature extraction and matching techniques for the face profile and ear, for example using deep learning algorithms; and (iv) testing fusion performance at rank and decision levels.

References

1. Mordini E, Tzovaras D (2012) Second generation biometrics: the ethical, legal and social context. Springer
2. Jain AK, Li SZ (2005) Handbook of face recognition. Springer, New York
3. Abaza A, Ross A, Herbert C, Harrison MF, Nixon MS (2013) A survey on ear biometrics. ACM Comput Surv 45:22:1–22:35
4. Santemiz P, Spreeuwers LJ, Veldhuis RNJ (2010) Side-view face recognition
5. Gentile JE, Bowyer KW, Flynn PJ (2008) Profile face detection a subset multi-biometric approach. In: Biometrics theory, applications and systems (BTAS)
6. Holz C, Buthpitiya S, Knaust M (2015) Bodyprint: biometric user identification on mobile devices using the capacitive touchscreen to scan body parts. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems, pp 3011–3014
7. Bicego M, Lagorio A, Grosso E, Tistarelli M (2006) On the use of SIFT features for face authentication. In: Computer vision and pattern recognition workshop
8. Dreuw P, Steingrube P, Hanselmann H, Ney H, Aachen G (2009) SURF-Face face recognition under viewpoint consistency constraints. In: BMVC
9. Ojala T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: application to face recognition. IEEE Trans Pattern Anal Mach Intell (PAMI) 28:2037–2041
10. Tan X, Triggs B (2007) Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: IEEE international workshop on analysis and modeling of faces and gestures (AMFG)
11. USTB. In: University of Science and Technology Beijing USTB Database. http://www1.ustb.edu.cn/resb/en/index.htm
12. UND. In: University of Notre Dame UND Databases. http://www3.nd.edu/~cvrl/CVRL/Data_Sets.html
13. Phillips PJ, Moon H, Rizvi SA, Rauss PJ (2000) The FERET evaluation methodology for face recognition algorithms. IEEE Trans Pattern Anal Mach Intell (PAMI) 22:1090–1104
14. Phillips PJ, Wechsler H, Huang J, Rauss PJ (1998) The FERET database and evaluation procedure for face recognition algorithms. Image Vis Comput (IVCJ) 16:295–306
15. Fahmy G, Elsherbeeny A, Mandala S, AbdelMottaleb M, Ammar H (2006) The effect of lighting direction/condition on the performance of face recognition algorithms. In: SPIE conference on human identification

16. Gross R, Baker S, Matthews I, Kanade T (2005) Handbook of face recognition: face recognition across pose and illumination. Springer
17. Zhang X, Gao Y (2009) Face recognition across pose: a review. Pattern Recogn 42:2876–2896
18. Ghaffary K, Tab F, Danyali H (2011) Profile-based face recognition using the outline curve of the profile Silhouette. In: IJCA special issue on Artificial intelligence techniques—Novel approaches and practical applications, pp 38–43
19. Kaufman GJ, Breeding KJ (1976) The automatic recognition of human faces from profile Silhouettes. IEEE Trans Syst Man Cybern SMC-6:113–121
20. Liposcak Z, Loncaric S (1999) A scale-space approach to face recognition from profiles. In: The 8th international conference on computer analysis of images and patterns (CAIP)
21. Harmon LD, Khan MK, Larsch R, Raming PF (1981) Machine identification of human faces. Pattern Recogn 13:97–110
22. Wu C, Huang J (1990) Human face profile recognition by computer. Pattern Recogn 23:255–259
23. Bir B, Xiaoli Z (2004) Face recognition from face profile using dynamic time warping. In: International conference on pattern recognition (ICPR)
24. Zhou X, Bhanu B (2005) Human recognition based on face profiles in video. In: IEEE computer society conference on computer vision and pattern recognition (CVPR)
25. Sihao D, Qiang Z, Yuan FZ, Dong X (2013) Side-view face authentication based on wavelet and random forest with subsets. In: IEEE international conference on intelligence and security informatics (ISI)
26. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444
27. Krizhevsky A, Sutskever I, Hinton G (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 1097–1105
28. Taigman Y, Yang M, Ranzato M, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. In: The IEEE conference on computer vision and pattern recognition, pp 1701–1708
29. Schroff F, Kalenichenko D, Philbin J (2015) Facenet: a unified embedding for face recognition and clustering. In: The IEEE conference on computer vision and pattern recognition, pp 815–823
30. Parkhi O, Vedaldi A, Zisserman A (2015) Deep Face Recogn BMVC 1:6–18
31. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: ICLR
32. Zhang Y, Mu Z (2017) Ear detection under uncontrolled conditions with multiple scale faster region-based convolutional neural networks. Symmetry 9:53–72
33. Chang K, Bowyer K, Sarkar S, Victor B (2003) Comparison and combination of ear and face images in appearance-based biometrics. IEEE Trans Pattern Anal Mach Intell (PAMI) 25:1160–1165
34. Kusuma D, Takashi Y (2006) Ear photo recognition using scale invariant keypoints. In: The 2nd international association of science and technology for development (IASTED)
35. Lowe D (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis (IJCV) 60:91–110
36. Kisku D, Mehrotra H, Gupta P, Sing J (2009) SIFT-based ear recognition by fusion of detected key-points from color similarity slice regions. In: The international conference on advances in computational tools for engineering applications (ACTEA)
37. Feng J, Mu Z (2009) Texture analysis for ear recognition using local feature descriptor and transform filter. In: SPIE pattern recognition and computer vision
38. Wang Y, Mu Z-C, Zeng H (2008) Block-based and multi-resolution methods for ear recognition using wavelet transform and uniform local binary patterns. In: The 19th international conference on pattern recognition (ICPR)
39. Yuan L, Mu Z, Liu Y (2006) Multimodal recognition using face profile and ear. In: The 1st international conference on systems and control in aerospace and astronautics (ISSCAA)
40. Xu X, Mu Z (2007) Multimodal recognition based on fusion of ear and profile face. In: The 4th international conference on image and graphics (ICIG)

41. Pan X, Cao Y, Xu X, Lu Y, Zhao Y (2008) Ear and face based multimodal recognition based on KFDA. In: International conference on audio, language and image processing (ICALIP)
42. Rahman MM, Ishikawa S (2005) Proposing a passive biometric system for robotic vision. In: The 10th international symposium on artificial life and robotics (AROB)
43. Rathore R, Prakash S, Gupta P (2013) Efficient human recognition system using ear and profile face. In: Biometrics theory, applications and systems (BTAS)
44. Abaza A, Hebert C, Harrison MF (2010) Fast learning ear detection for real-time surveillance. In: Fourth IEEE international conference on biometrics: theory applications and systems (BTAS)
45. Lowe D (1999) Object recognition from local scale-invariant features. In: IEEE international conference on computer vision (ICCV)
46. Wang J-G, Li J, Lee CY, Yau W-Y (2010) Dense SIFT and Gabor descriptors-based face representation with applications to gender recognition. In: IEEE international conference on control, automation, robotics and vision (ICARCV)
47. Bay H, Ess A, Tuytelaars T, Gool LV (2008) SURF speeded up robust features. Comput Vis Image Underst (CVIU) 110:346–359
48. Ojala T, Pietikinen M, Harwood D (1996) A comparative study of texture measures with classification based on featured distributions. Pattern Recogn 29:51–59
49. DeCann B, Ross A (2013) Relating roc and cmc curves via the biometric menagerie. In: Biometrics theory, applications and systems (BTAS)
50. Jain AK, Flynn P, Ross A (2007) Handbook of biometrics: introduction to biometrics. Springer

Chapter 5
Cross-Spectral Iris Matching for Surveillance Applications

Mohammed A. M. Abdullah, Raid R. Al-Nima, Satnam S. Dlay, Wai L. Woo and Jonathon A. Chambers

Abstract With the advancement in iris recognition at a distance, cross-spectral iris matching is emerging as a hot topic. The importance of cross-spectral matching stems from the feasibility of performing matching in several security applications such as watch-list identification, security surveillance and hazard assessment. Typically, a person's iris images are captured under Near-Infrared light (NIR) but most of the security cameras operate in the Visible Light (VL) spectrum. In this work, we therefore propose two methods for cross-spectral iris recognition capable of matching iris images in different lighting conditions. The first method is designed to work with registered iris images. The key idea is to synthesize the corresponding NIR images from the VL images using Artificial Neural Networks (ANN). The second one is capable of working with unregistered iris images based on integrating the Gabor filter with different photometric normalization models and descriptors along with decision level fusion to achieve the cross-spectral matching. Experimental and comparative results on the UTIRIS and the PolyU databases demonstrate that the proposed methods achieve promising results. In addition, the results indicate that the VL and NIR images provide complementary features for the iris pattern and their fusion improves the recognition performance.

Keywords Iris recognition · Cross-spectral matching · Multi-spectral recognition
Neural network · Photometric normalization · Surveillance at-a-distance

M.A.M Abdullah (✉) · R.R. Al-Nima · S.S. Dlay · W.L. Woo · J.A. Chambers
School of Electrical and Electronic Engineering, Newcastle University,
Newcastle upon Tyne, UK
e-mail: m.a.m.abdullah@ncl.ac.uk

R.R. Al-Nima
e-mail: r.r.o.al-nima@ncl.ac.uk

S.S. Dlay
e-mail: satnam.dlay@ncl.ac.uk

W.L. Woo
e-mail: lok.woo@ncl.ac.uk

J.A. Chambers
e-mail: Jonathon.Chambers@ncl.ac.uk

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_5

Introduction

Among the various traits used for human identification, the iris pattern has gained an increasing amount of attention for its accuracy, reliability, and noninvasive characteristics. In addition, iris patterns possess a high degree of randomness and uniqueness which is true even between identical twins and the iris remains stable throughout an adult's life [1, 2].

The initial pioneering work on iris recognition, which is the basis of many functioning commercial systems, was conducted by Daugman [1]. The performance of iris recognition systems is impressive as demonstrated by Daugman [3] who reported false acceptance rates of only 10^{-6} on a study of 200 billion cross-comparisons. Additionally, the potential of iris biometrics has also been affirmed with 1.2 trillion comparison tests carried out by the National Institute of Standards and Technology (NIST) which confirmed that iris biometrics has the best balance in terms of accuracy, template size and speed compared to other biometric traits [4].

Iris recognition technology is nowadays widely deployed in various large scale applications such as the border crossing system in the United Arab Emirates, Mexico national ID program and the Unique Identification Authority of India (UIDAI) project [5]. As a case in point, more than one billion residents have been enrolled in the UIDAI project where about 10^{15} all-to-all check operations are carried out daily for identity de-duplication using iris biometrics as the main modality [5, 6].

Nearly all currently deployed iris recognition systems operate predominately in the Near-Infrared (NIR) spectrum capturing images at 800–900 nm wavelength. This is because there are fewer reflections coming from the cornea and the dark pigmented irides look clearer under NIR light. In addition, external factors such as shadows and diffuse reflections become less under NIR light [7, 8]. Nevertheless, research in VL iris recognition has been gaining more attention in recent years due to the interest in iris recognition at a distance [9, 10]. In addition, competitions such as the Noisy Iris Challenge Evaluation (NICE) [11] and the Mobile Iris Challenge Evaluation [12] focus on the processing of VL iris images. This attention to visible wavelength-based iris recognition is boosted by several factors such as: (1) visible range cameras can acquire images from long distance and they are cheaper than NIR cameras; (2) surveillance systems work in the visible range by capturing images of the body, face and iris which could be used later for authentication [13].

Cross-domain matching can be divided into two main categories: cross-sensor matching and cross-spectral matching. Cross-sensor matching represents the case when the iris images acquired from one iris sensor must be matched against another sensor. This cross-sensor matching is believed to degrade recognition performance due to several factors such as the variations in lenses, sensor sensitivity and the difference in NIR illumination [14]. On the other hand, cross-spectral matching represents the case of matching iris images taken under different illumination conditions i.e. NIR versus VL. Cross-spectral matching is a challenging task because there are considerable differences among images pertaining to different wavelength bands. As the topic of this chapter is surveillance and security applications, the focus will be on the second type (cross-spectral matching).

Motivations and Contributions

Related to the aforementioned problem, since both VL and NIR iris recognition systems are now widely deployed, cross-spectral matching is paramount due to its use in several security applications such as surveillance at-a-distance and automated watch-list identification. Typically, a person's iris images are captured under NIR but most security cameras operate in the VL spectrum. Hence, NIR versus VL matching is desired. In addition, studying the performance difference of iris recognition systems exploiting NIR and VL images is important because it gives insight into the essential features in each wavelength which in turn helps to develop a robust automatic identification system.

In this chapter, we propose two methods for the problem of VL to NIR iris matching (and vice versa) dealing with unregistered and registered iris images belonging to the same subject. In addition, we investigate the difference in iris recognition performance with NIR and VL imaging. In particular, we investigate iris performance in each channel (red, green, blue and NIR) and the feasibility of cross-channel authentication (i.e. NIR vs. VL). Furthermore, enhancing the iris recognition performance with multi-channel fusion is achieved.

In summary, the key contribution of this chapter can be summarized as follows:

- A framework for cross-spectral iris recognition capable of matching registered and unregistered iris images captured under different lighting conditions.
- Filling the gap in multi-spectral iris recognition by exploring the performance difference in iris biometrics under NIR and VL imaging.
- Boosting iris recognition performance with multi-channel fusion.

The rest of this chapter is organized as follows: related work is given in Section "Related Work". The proposed framework for cross-spectral iris matching is explained in Section "Cross-Spectral Iris Matching Approaches". The "Results and Discussion" section presents experimental results and the discussion, while "Conclusions" section concludes this chapter.

Related Work

Iris recognition technology has witnessed a rapid development over the last decade driven by its wide applications in the world. At the outset, Daugman [1] proposed the first working iris recognition system which has been adopted later by several commercial companies such as IBM, Irdian and Oki. Several works followed after that, but almost all of them assessed iris recognition performance under NIR.

In terms of classification with deep learning, Du et al. [15] developed a method to automatically classify left versus right iris images using a convolutional neural network. The method can be adapted to different databases and is not limited to the occlusion of the tear-duct or the lateral angle, achieving a classification performance of 97.5%.

Minae et al. [16] investigated the application of the deep features extracted using the VGG-Net convolutional network within iris recognition. The proposed method is tested on the CASIA-1000 and ITT iris databases with an accuracy of 99.4%. Similarly, Raja et al. [17] proposed multi-patch deep feature extraction using deep sparse filters to form a reliable smartphone iris verification system using the visible spectrum. The method is evaluated on the MICHE-I dataset and the authors reported an Equal Error Rate (EER) less than 2%.

The demand for more accurate and robust biometric systems has increased with the expanded deployment of large-scale national identity programs. Hence, researchers have investigated iris recognition performance under different wavelengths; or the possibility of fusing NIR and VL iris images to enhance recognition performance. Nevertheless, inspecting the correlation between NIR and VL iris images has been understudied and the problem of cross-spectral iris recognition is still unsolved.

Boyce et al. [18] explored iris recognition performance under different wavelengths on a small multi-spectral iris databases consisting of 120 images from 24 subjects. According to the authors, higher accuracy was achieved for the red channel compared to green and blue channels. The study also suggested that cross-channel matching is feasible. The iris images were fully registered and captured under ideal conditions. In [19] the authors employed the feature fusion approach to enhance the recognition performance of iris images captured under both VL and NIR. The wavelet transform and discrete cosine transform were used for feature extraction while the features were augmented with the ordered weighted average method to enhance the performance.

In Ngo et al. [20] a multi-spectral iris recognition system was implemented which employed eight wavelengths ranges from 405–1550 nm. The results on a database of 392 iris images showed that the best performance was achieved with a wavelength of 800 nm. Cross-spectral experimental results demonstrated that the performance degraded with larger wavelength difference. Ross et al. [21] explored the performance of iris recognition in wavelengths beyond 900 nm. In their experiments, they investigated the possibility of observing different iris structures under different wavelengths and the potential of performing multi-spectral fusion for enhancing iris recognition performance. Similarly, Ives et al. [22] examined the performance of iris recognition under a wide range of wavelengths between 405–1070 nm. The study suggests that illumination wavelength has a significant effect on iris recognition performance. Hosseini et al. [8] proposed a feature extraction method for iris images taken under VL using a shape analysis method. Potential improvement in recognition performance was reported when combining features from both NIR and VL iris images taken from the same subject.

Recently, a method has been proposed in [23] for cross-spectral periocular verification using an Artificial Neural Network (ANN). Experiments were conducted on a database consisting of eye images captured under VL, night vision and NIR. Alonso-Fernandez et al. [24] conducted comparisons on the iris and periocular modalities and their fusion under NIR and VL imaging. However, the images were not taken from the same subjects as the experiments were carried out on different databases (three databases contained close-up NIR images, and two others contained VL images). Unfortunately, this may not give an accurate indication about the iris performance as the images do not belong to the same subject. In [25] the authors suggested enhancing iris recognition performance in non-frontal images through multi-spectral fusion of the iris pattern and scleral texture. Since the scleral texture is better seen in VL and the iris pattern is observed in NIR, multi-spectral fusion could improve the overall performance.

In terms of cross-spectral iris matching, the authors in [13] proposed an adaptive method to predict the NIR channel image from VL iris images using an ANN. Similarly, Burge and Monac [26, 27] proposed a model to predict NIR iris images using features derived from the color and structure of the VL iris images.

In our previous work [28] we explored the differences in iris recognition performance across the VL and NIR spectra. In addition, we investigated the possibility of cross-channel matching between the VL and NIR imaging. The cross-spectral matching turns out to be challenging with an EER larger than 27% in the UTIRIS database. Lately, Ramaiah and Kumar [29] emphasized the need for cross-spectral iris recognition and introduced a database of registered iris images and conducted experiments on iris recognition performance under both NIR and VL. The results of cross-spectral matching achieved an EER larger than 34% which confirms the challenge of cross-spectral matching. The authors concluded their paper by: "it is reasonable to argue that cross-spectral iris matching seriously degrades the iris matching accuracy".

More recently, Ramaiah and Kumar [14] introduced a new iris database containing NIR and VL iris images acquired simultaneously and performed experiments to improve the cross-sensor and cross-spectral iris matching. In our previous work [30], we proposed a framework for cross-spectral iris matching capable of working with unregistered iris images pertaining to the same subject. Experiments on the UTIRIS database showed that the proposed framework notably improved the cross-spectral matching performance. In this chapter, this framework is further investigated and extended to work with registered iris images to cover all the possible conditions.

Cross-Spectral Iris Matching Approaches

Matching across iris images captured in VL and NIR is a challenging task because there are considerable differences among such images pertaining to different wavelength bands. Although the appearance of different spectrum iris images is not constant, the structure is the same as they belong to the same person. Generally speaking, to alleviate the perceptual differences, two approaches can be adopted: a training-based approach and a photometric/descriptor-based approach. The training-based approach is known to achieve outstanding performance for cross-spectral matching of registered images [14, 23] but the training-based approach tends to fail due to the lack of pixel correspondence among unregistered images.

In this section, two methods are proposed for cross-spectral iris matching. The first one is designed to work with registered iris images in the verification mode.

The key idea is to synthesize the corresponding NIR images from the VL images using the ANN techniques. After that, the predicted NIR images can be matched against the VL images to perform cross-spectral matching. The second method is capable of working with unregistered iris images in the identification mode based on integrating the Gabor filter with different photometric normalization models and descriptors along with decision level fusion. Details of these methods are described in the next sections.

Matching of Registered Images

The problem of training a predictive model is considered to be a non-linear multivariate regression problem. As cross-spectral matching can be stated to be a non-linear matching, the predictive model is estimated using the ANN technique due to the latter ability in establishing a non-linear mapping between inputs and targets [31]. Accordingly, after estimating the mapping, the target parameters (NIR pixels) are predicted from the input parameters (gray-scale pixels of a query image).

A Feed Forward Neural Network (FFNN) is used to build the predictive model for cross-spectral matching. It is trained on the gray-scale pixels of the color iris images and the target is set to be the corresponding NIR pixels. The training data which consists of 30 classes, are randomly divided into two subsets. The learning subset consists of 60% of the training data while the testing subset is composed of the remaining 40%. In our experiments, the FFNN is composed of two hidden-layers. The number of layers and neurons is chosen as a trade-off between complexity and network performance. The iris images are segmented and normalized as proposed in our previous work [32] to ensure only iris pixels are fed to the network.

In the training phase, training is performed using the scaled conjugate gradient backpropagation approach with a transfer function of tangent sigmoid in the hidden layer and a linear activation function in the output layer then the final weights are stored for each subject. The training process stops when the mean square error equals to 10^{-6}. During the testing phase, these weights are used to generate the predicted NIR iris image. The block diagrams of the training and testing phase are illustrated in Fig. 5.1.

Matching of Unregistered Images

Cross-spectral matching of unregistered images is more challenging due to lack of pixels correspondence in the unregistered images. Therefore, more sophisticated methods are needed to address the spectral bands and illumination variations.

We have exploited various photometric normalization techniques and descriptors to alleviate these differences. In this context, we employed the Binarized Statistical Image Features (BSIF) descriptor [33], Difference-of-Gaussian (DoG) filtering in

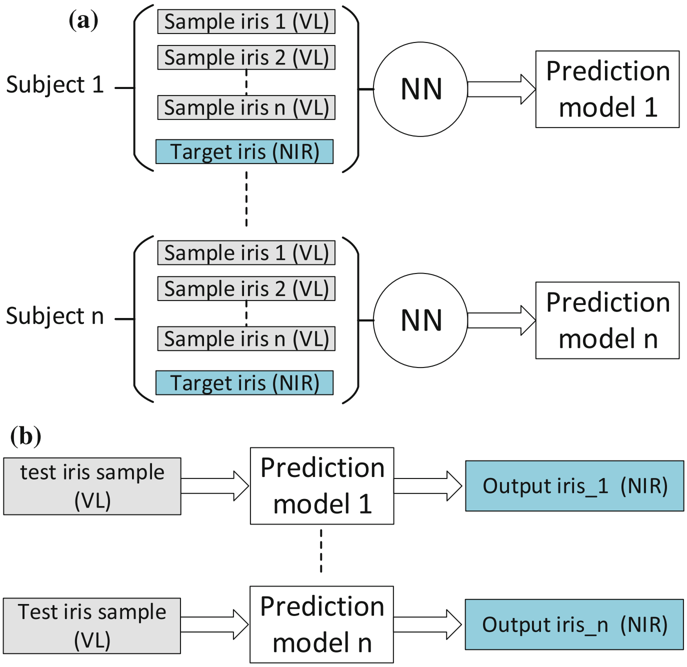


Fig. 5.1 Block diagram of the prediction model: **a** training phase **b** testing phase

addition to a collection of the photometric normalization techniques available from the INface Toolbox[1] [34, 35]: adaptive single scale retinex, non-local means, wavelet based normalization, homomorphic filtering, multi-scale quotient, Tan and Triggs normalization and Multi Scale Weberface (MSW). Among these illumination techniques and descriptors, the DoG, BSIF and MSW are found to reduce the iris cross-spectral variations. The reader can find more about these methods in our previous work [30].

Extensive experiments demonstrated that using one of the aforementioned photometric normalization methods alone is not sufficient to achieve an acceptable iris recognition performance for unregestered iris images with EER $> 17\%$. Therefore, we propose to integrate the Gabor filter with these methods in addition to decision level fusion to achieve a robust cross-spectral iris recognition. Also, using the phase information of the Gabor filter rather than amplitude is known to result in robustness to different variations such as: illumination variations, imaging contrast and camera gain [7].

Hence, we propose to integrate the 1D log-Gabor filter [36] with DoG, BSIF and MSW to produce the G-DoG, G-BSIF and G-MSW (where G stands for Gabor) in addition to decision level fusion to achieve a robust cross-spectral iris recognition. The block diagram of the proposed framework is depicted in Fig. 5.2.

[1] http://luks.fe.uni-lj.si/sl/osebje/vitomir/face_tools/INFace/.

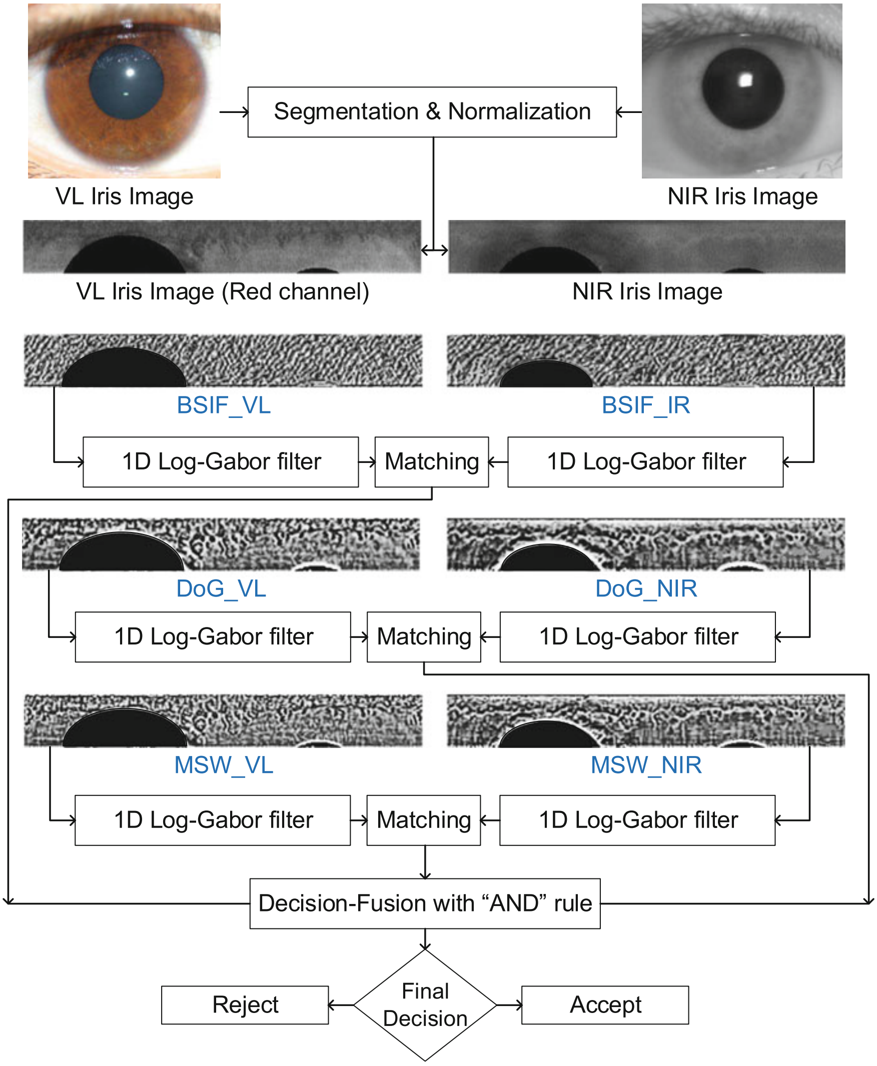


Fig. 5.2 Block diagram of the proposed cross-spectral matching framework

Results and Discussion

In this work our aim is to ascertain true cross-spectral iris matching using images taken from the same subject under VL and NIR spectra. In addition, we investigate the iris biometric performance under different imaging conditions and the fusion of VL+NIR images to boost the recognition performance. The recognition performance is measured with the EER and the Receiver Operating Characteristic (ROC) curves.

Databases

The cross-spectral experiments are conducted using two publicly available cross-spectral iris databases, the PolyU bi-spectral iris database [14] and the UTIRIS database [8]. These databases are briefly described next.

The PolyU iris database contains bi-spectral iris images which are captured under NIR and VL simultaneously from both right and left eyes of 209 subjects. Each image has a resolution of 640×480 and available in both NIR and VL with pixel correspondences. The UTIRIS database contains two sessions with 1540 images; the first session was captured under VL while the second session was captured under NIR. Each session has 770 images taken from the left and right eye of 79 subjects where each subject has an average of 5 iris images.

Pre-processing and Feature Extraction

Typically, an iris recognition system operates by extracting and comparing the pattern of the iris in the eye image. These operations involve four main steps namely: image acquisition, iris segmentation, normalization, feature extraction and matching [7].

All iris images were segmented and normalized using the robust iris segmentation algorithm from our previous work [32]. For the PolyU database, the NIR images were segmented first and the same parameters were used to segment the corresponding VL images because the images are registered. On the other hand, for the UTIRIS database, it is noticed that the red channel gives the best segmentation results because the pupil region in this channel contains the smallest amount of reflection as shown in Figs. 5.3 and 5.4. The images in the VL session were down-sampled by two in each dimension to obtain the same size as the images in the NIR session.

For feature extraction, the normalized iris image is convolved with the 1D log-Gabor filter to extract the features where the output of the filter is phase quantized to four levels to form the binary iris vector [36]. After that, the Hamming distance is used to find the similarity between two IrisCodes in order to decide if the vectors belong to the same person or not.

Light-Eyed versus Dark-Eyed

Capturing iris images under NIR light eliminates most of the rich melanin information because the chromophore of the human iris is only visible under VL [8, 37]. Therefore, light pigmented irides exhibit more information under visible light. Figure 5.3 shows a green-yellow iris image captured under NIR and VL. It can be seen that the red channel reveals more information than the NIR image. So,

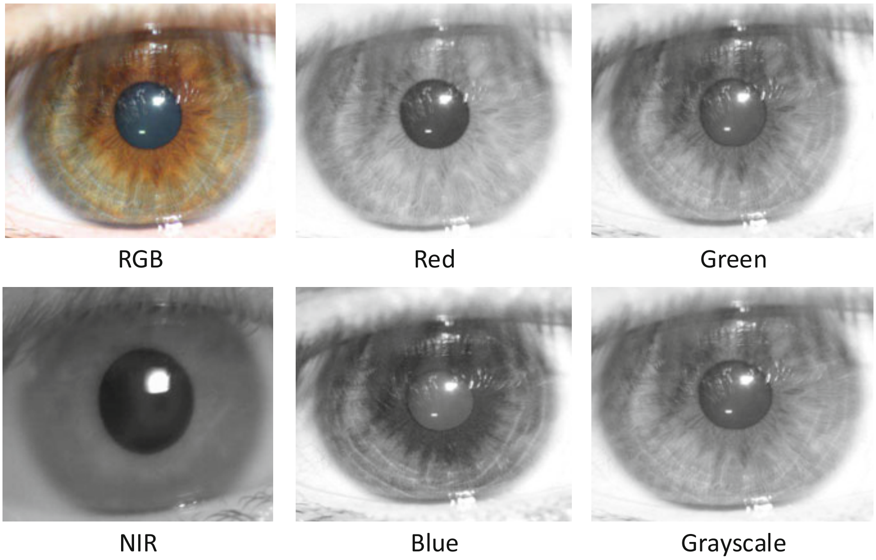


Fig. 5.3 Green-yellow iris image decomposed into red, green, blue and grayscale with the NIR counterpart

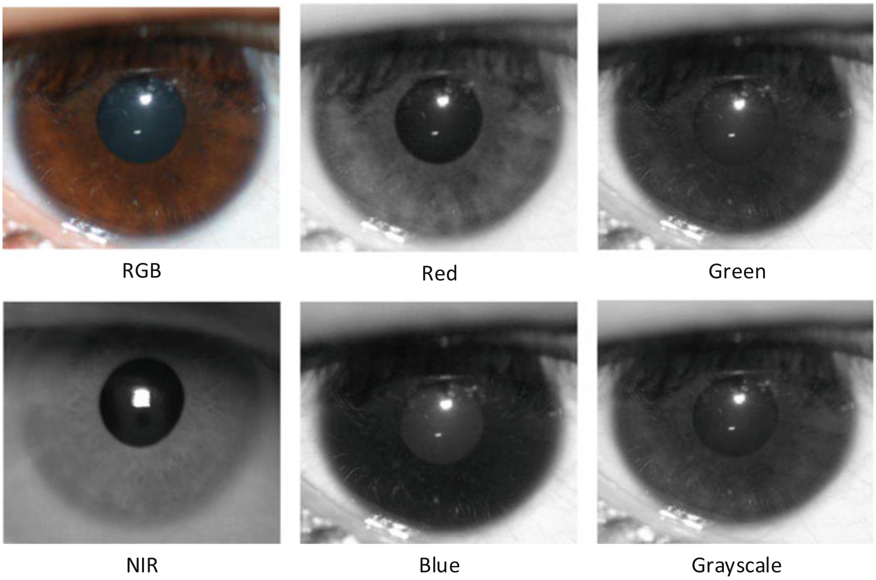


Fig. 5.4 Brown iris image decomposed into red, green, blue and grayscale with the NIR counterpart

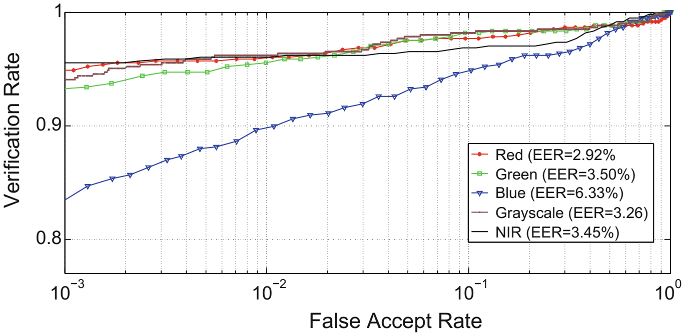


Fig. 5.5 The performance of the iris recognition under red, green, blue and NIR spectra for the UTIRIS database

intuitively the recognition performance would be better for such images in the VL rather than the NIR spectrum.

On the other hand, with dark pigmented irides, stromal features of the iris are only revealed under NIR and they become hidden in VL so the information related to the texture is revealed rather than the pigmentation as shown in Fig. 5.4. Therefore, the recognition performance for the dark pigmented irides would give better results if the images were captured under NIR spectrum.

We carried out experiments on each channel (i.e. NIR, red, green and blue) and measured the performance using ROC and EER. For the UTIRIS database, it can be seen from Fig. 5.5 that the best performance is achieved under the red channel with EER = 2.92% followed by the green channel with EER = 3.50% while the blue channel achieved worse results with EER = 6.33%. It is also noticed that NIR images did not give the best performance for this database (EER = 3.45%). This is because most of the iris images in the UTIRIS database are light pigmented.

On the contrary, as most of the images in the PolyU iris database are dark pigmented, the NIR images achieved the best results as shown in Fig. 5.6 with EER = 2.71% while the red channel achieved a worse EER of 7.49% followed by the green and blue channels with EER of 9.09% and 16.02%, respectively.

Cross-Spectral Experiments

Cross-spectral study is important because it shows the feasibility of performing iris recognition in several security applications such as information forensics, security surveillance and hazard assessment. Typically, a person's iris images are captured under NIR but most of the security cameras operate in the VL spectrum. Hence, NIR versus VL matching is desired.

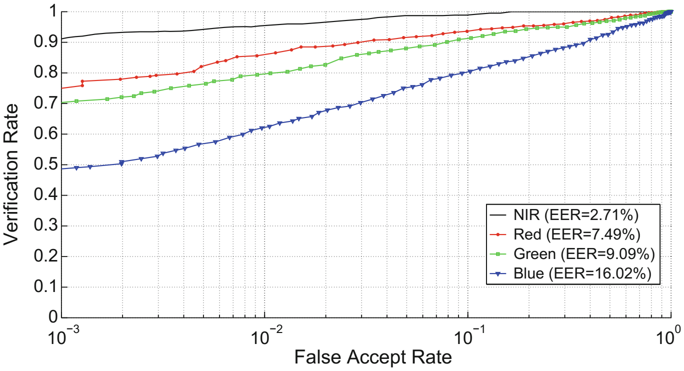


Fig. 5.6 The performance of the iris recognition under red, green, blue and NIR spectra for the PolyU database

Table 5.1 EER (%) of different channels comparison on the UTIRIS database

	NIR	Red	Green	Blue
NIR	3.45	27.53	38.81	40.31
Red	–	**2.92**	3.64	15.34
Green	–	–	3.50	6.45
Blue	–	–	–	6.33

Table 5.2 EER (%) of different channels comparison on the PolyU database

	NIR	Red	Green	Blue
NIR	**2.71**	17.16	22.74	36.17
Red	–	7.49	10.37	14.34
Green	–	–	9.09	13.48
Blue	–	–	–	16.02

In this context, we carried out these comparisons using the traditional 1D Log-Gabor filter: NIR versus red, NIR versus green and NIR versus blue on the UTIRIS and PolyU databases.

Figures 5.5 and 5.6 depict the ROC curves of these comparisons for the UTIRIS database and the PolyU database, respectively. Accordingly, the green and blue channels resulted in bad performance due to the big gap in the electromagnetic spectrum between these channels and the NIR spectrum.

On the contrary, the red channel gave the best performance compared to the green and blue channels. This can be attributed to the small gap in the wavelength of the red channel (780 nm) compared to the NIR (850 nm). Therefore, the comparisons of Red vs NIR is considered as the baseline for cross-spectral matching. Tables 5.1 and 5.2 show the EERs of cross-channel matching experiments for both the UTIRIS and PolyU databases, respectively.

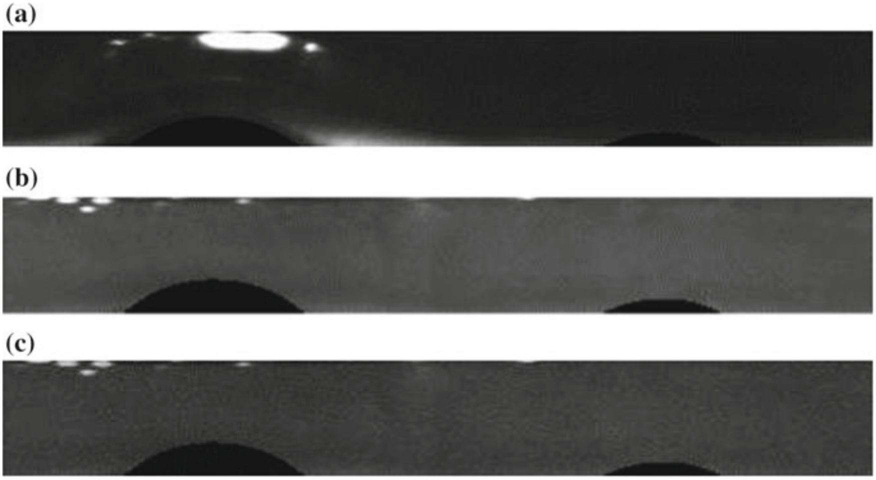


Fig. 5.7 Unwrapped iris images: **a** R channel, **b** NIR image and **c** the predicted NIR image

Table 5.3 The results of the cross-spectral verification using the FFNN on the PolyU database

Matching	EER
NIR versus NIR	2.71
NIR versus NIR_Predicted	2.75
NIR versus Red	17.16

For all cross-spectral experiments, we have adopted the leave-one-out approach to obtain the comparison results [38]. Hence, for each subject with (m) iris samples, we have set one sample as a probe and the comparison is repeated iteratively by swapping the probe with the remaining ($m - 1$) samples. The experiments for each subject are repeated ($m(m - 1)/2$) times and the final performance is measured in terms of EER by taking the minimum of the obtained comparison scores of each subject.

Cross-Spectral Matching of Registered Images

A training based approach is adopted for the iris images in the PolyU database as these images are having the same pixel correspondence. An FFNN is built for each subject and the unwrapped iris images of the same subject were converted to the gray-scale and used as an input to the ANN.

The target of the ANN is set to be one of the corresponding NIR images of the same subject. In the testing phase, the output image (NIR pixels) is predicted from the input VL image and the verification is performed. Figure 5.7 shows the input VL image and the predicted NIR image while Table 5.3 shows the cross-spectral recognition performance.

Cross-Spectral Matching of Unregistered Images

The cross-spectral matching is more challenging for unregistered images as the case in the UTIRIS database and training based approach tend to fail on such images due to the lack in pixel correspondence. Therefore, to alleviate the cross-spectrum differences, different descriptors and feature enhancement techniques are employed, out of which the DoG, MWS and BSIF recorded the best results as shown in Table 5.4. Hence, the proposed framework, which is depicted in Fig. 5.2, is based on these descriptors.

To further enhance the performance of cross-spectral matching, the fusion of the G-DoG, G-BSIF and G-MSW is considered. Different fusion method are investigated with the UTIRIS database namely: feature fusion, score fusion and decision fusion; out of which the decision fusion is observed to be the most effective.

Table 5.5 shows the performance of different fusion strategies for cross-spectral matching in terms of EER. Feature fusion resulted in poor results where the EER varied from 14–18%. Score level fusion with minimum rule achieved better results.

Table 5.4 Experiments on different descriptors for cross-spectral matching on the UTIRIS database

Method	EER (%)
Baseline	27.53
LBP (different combinations) [39]	>28
Adaptive single scale retinex	25.56
Non-local means normalization	27.49
Wavelet based normalization	28.65
Homomorphic filtering	29.07
Multi scale self quotient	26.99
Tan and Triggs normalization	23.43
DoG	**19.51**
MSW	**18.91**
BSIF	**20.64**

Table 5.5 Experiments on different fusion strategies for cross-spectral matching on the UTIRIS database

	EER (%)		
Method	Feature fusion	Score fusion (min)	Decision fusion (AND)
DoG + MSW	16.56	14.21	8.08
DoG + BSIF	17.56	15.42	8.77
BSIF + MSW	18.12	16.34	8.33
DoG + BSIF + MSW	14.59	12.83	**6.81**

On the other hand, AND rule decision level fusion achieved the best results with EER = 6.81%.

A low False Accept Rate (FAR) is preferred to achieve a secure biometric system. To enhance the performance of our system and reduce the FAR, a fusion at the decision level is performed. Therefore, the conjunction "AND" rule is used to combine the decisions from the G-DoG, G-BSIF and G-MSW. This means a false accept can only happen when all the previous descriptors produce a false accept [40].

Let $PD(FA)$, $PS(FA)$ and $PM(FA)$ represent the probability of a false accept using G-DoG, G-BSIF and G-MSW respectively. Similarly, $PD(FR)$, $PS(FR)$ and $PM(FR)$ represent the probability of a false reject. Therefore, the combined probability of a false accept $PC(FA)$ is the product of the three probabilities of the descriptors:

$$PC(FA) = PD(FA).PS(FA).PM(FA). \tag{1}$$

On the other hand, the combined probability of a false reject $PC(FR)$ can be expressed as the complement of the probability that none of the descriptors produce a false reject:

$$\begin{aligned} PC(FR) &= (PD(FR)'.PS(FR)'.PM(FR)')', \\ &= (1 - (1 - PD(FR))(1 - PS(FR))(1 - PM(FR))), \\ &= PD(FR) + PS(FR) + PM(FR) + \\ &PD(FR).PS(FR) + PD(FR).PM(FR) + \\ &PS(FR).PM(FR) + PD(FR).PS(FR).PM(FR). \quad (5) \end{aligned}$$

It can be seen from the previous equations that the joint probability of false rejection increases while the joint probability of false acceptance decreases when using the AND conjunction rule.

The proposed framework for cross-spectral matching has also been applied on the PolyU iris database and an enhancement in the cross-spectral matching is recorded as shown in Fig. 5.9. However, this improvement is not as good as the performance improvement of the UTIRIS database as the images in the PolyU database are registered. The training based approach achieve better results however it might be difficult to acquire registered images in the real life situations.

Multi-spectral Iris Recognition

The cross-channel comparisons demonstrated that red and NIR channels are the most suitable candidates for fusion as they gave the lowest EER compared to other channels as shown in Figs. 5.8 and 5.9 so it is common sense to fuse them in order to boost the recognition performance. Score level fusion is adopted in this paper due to its efficiency and low complexity [41]. Hence, we combined the matching scores (Hamming distances) from both the red and NIR images using sum rule based fusion

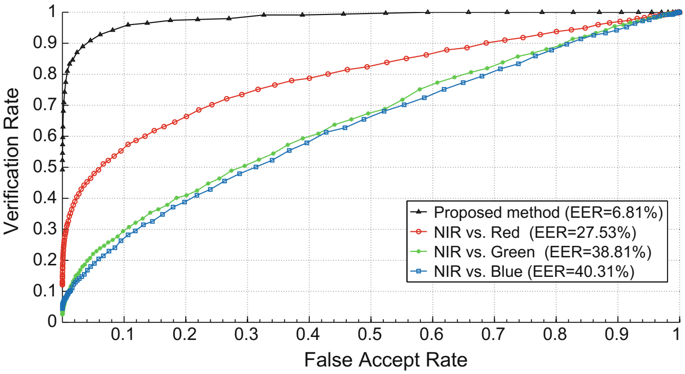


Fig. 5.8 Cross-channel matching of the UTIRIS database

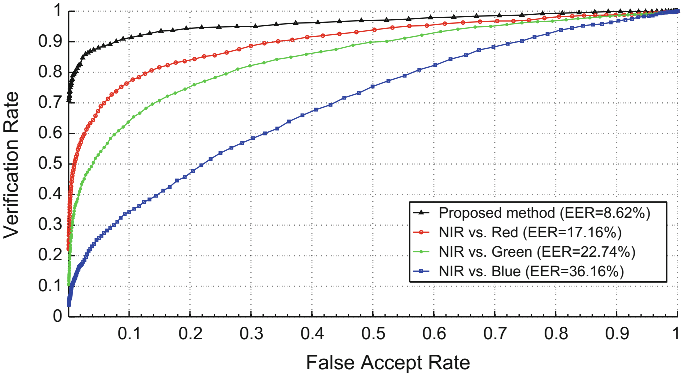


Fig. 5.9 Cross-channel matching of the PolyU database

with equal weights to generate a single matching score. After that the recognition performance is evaluated again with the ROC curves and EER.

The VL and NIR images in the UTIRIB database are not registered. Therefore, they provide different iris texture information. It is evident from Fig. 5.10 that such fusion is useful to the iris biometric as there is a significant improvement in the recognition performance after the fusion with EER of only 0.54% compared to 2.92 and 3.45% before the fusion.

On the other hand, the fusion of the VL and NIR images in the PolyU database barely improved the recognition performance (Fig. 5.11). This is because the iris images in the PolyU database are registered and hence they do not provide much different information as the iris texture is analogous.

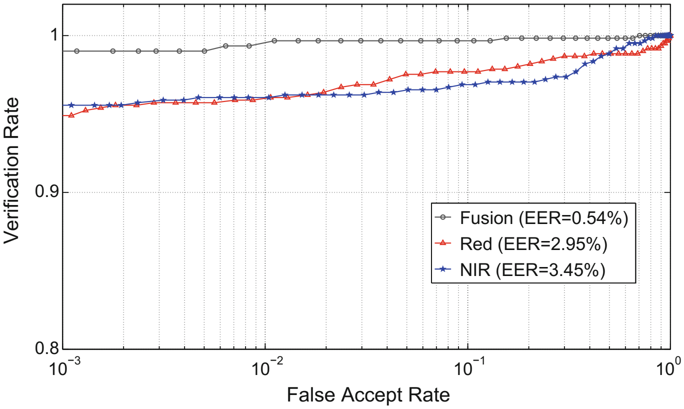


Fig. 5.10 ROC curves showing the iris recognition performance before and after fusing the information of the red and NIR channel for the UTIRIS database

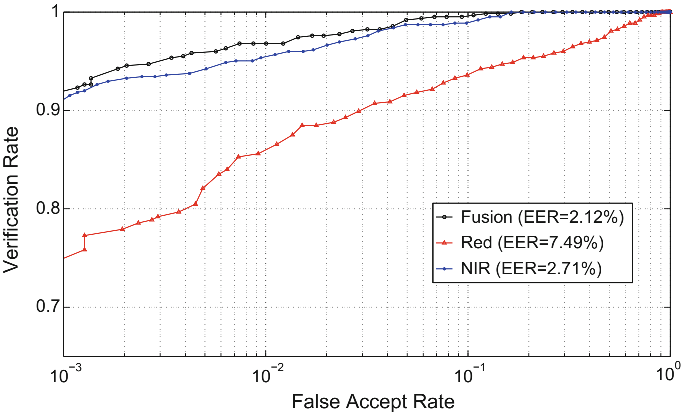


Fig. 5.11 ROC curves showing the iris recognition performance before and after fusing the information of the red and NIR channel for the PolyU database

Comparisons with Related Work

As mentioned earlier, the previous works have either adopted training models [13] to predict the corresponding iris images or descriptor models to alleviate the difference among iris images captured under different lighting conditions [30].

In [13] the authors adopted a training based approach using the FFNN to predict the NIR from the VL images. Although good results were reported, the main drawback is adopting a sequential process to train the NN. The problems of using this style are: long time in the training phase, high error values in the training and testing phases; and increasing the possibility of the local minima. On the contrary, constructing the output (NIR images) in one burst output will overcome the

Table 5.6 Cross-spectral matching comparison with different methods

Method	Database	EER (%)
Wild et al. [42]	UTIRIS	33–55
Ramaiah and Kumar [29]	Private	34.01
Abdullah et al. [28]	UTIRIS	27.46
Proposed	UTIRIS	6.81
Proposed	PolyU	2.75

aforementioned drawbacks. It can be argued that this style would increase the memory requirements, but this issue can be overlooked taking in mind the memory capacity of current computers.

In the works of [28, 42], the results of cross-spectral matching on unregistered iris images were reported. However, no models were proposed to enhance the cross-spectral iris matching. Table 5.6 shows the comparison results of the aforementioned works compared to our method.

Processing Time

All experiments were conducted on a 3.2 GHz core i5 PC with 8 GB of RAM under the Matlab environment. The average training time for the FFNN is 7.5 minutes while the testing time was less than two seconds.

On the other hand, the proposed framework for cross-spectral matching of unregistered images consists of four main descriptors namely: BSIF, DoG, MSW and 1D log-Gabor filter. The processing times of the 1D log-Gabor filter, BSIF and DoG descriptors are 10, 20 and 70 ms, respectively while the MSW processing times is 330 ms. Therefore, the total computations time of the proposed method is less than half a second which implies its suitability for real time applications.

Applicability and Limitations

The proposed method for cross-spectral matching of unregistered iris images can be applied directly without the need for the iris image pairs to be aligned. However, as mentioned earlier, the performance of a photometric/descriptor-based approach is not as good as the training-based approaches in terms of recognition performance.

On the other hand, the training-based approaches are known to achieve outstanding performance for cross-spectral matching of registered images but they the limitation of matching registered pairs only. Although it cannot be guaranteed that the iris pairs will be aligned when applying the proposed method for registered iris cross-matching, applying real-time image registration techniques [43, 44] will pave the way to utilize the proposed method in practical applications.

Therefore, for applications that require high performance, training-based approaches can be used along with image registration techniques. On the other hand, for less demanding applications, the proposed method for matching unregistered images can be applied readily without the need for image registration or train models.

Conclusions

In this chapter, a framework for cross-spectral iris matching was proposed. In addition, this work highlights the applications and benefits of using multi-spectral iris information in iris recognition systems. We investigated iris recognition performance under different imaging channels: red, green, blue and NIR. The experiments were carried out on the UTIRIS and the PolyU databases and the performance of the iris biometric was measured.

Two methods were proposed for cross-spectral iris matching. The first method was designed to work with registered iris images which adopted an FFNN to synthesize the corresponding NIR images from the VL images to perform the verification. The second method is capable of working with unregistered iris images based on integrating the Gabor filter with different photometric normalization models and descriptors along with decision level fusion to achieve the cross-spectral matching.

We drew the following conclusions from the results. According to Table 5.4, among a variety of descriptors, the Difference of Gaussian (DoG), Binarized Statistical Image Features (BSIF) and Multi-scale Weberface (MSW) were found to give good cross-spectral performance for unregistered iris images after integrating them with the 1D log-Gabor filter. Table 5.6 and Fig. 5.8 showed a significant improvement in the cross-spectral matching performance of the UTIRIS database using the proposed framework.

In terms of multi-spectral iris performance, Fig. 5.5 showed that the red channel iris images in the UTIRIS database achieved better performance compared to other channels or the NIR imaging. This can be attributed to the large number of the light pigmented irides in this database. However, the NIR images in the PolyU database achieved the best performance among other channels due to the domination of the brown irides in the PolyU database.

It was also noticed from Figs. 5.8 and 5.9 that the performance of the iris recognition varied as a function of the difference in wavelength among the image channels. Fusion of the iris images from the red and NIR channels improved the recognition performance in the UTIRIS database. The results implied that both VL and NIR imaging were important to form a robust iris recognition system as they provided complementary features for the iris pattern.

References

1. Daugman J (1993) High confidence visual recognition of persons by a test of statistical independence. IEEE Trans Pattern Anal Mach Intell 15(11):1148–1161
2. Sun Z, Tan T (2009) Ordinal measures for iris recognition. IEEE Trans Pattern Anal Mach Intell 31(12):2211–2226
3. Daugman J (2006) Probing the uniqueness and randomness of iriscodes: results from 200 billion iris pair comparisons. Proc IEEE 94(11):1927–1935
4. Grother PJ, Quinn GW, Matey JR, Ngan ML, Salamon WJ, Fiumara GP, Watson CI (2012) IREX III: performance of iris identification algorithms. Report, National Institute of Standards and Technology
5. Jain AK, Nandakumar K, Ross A (2016) 50 years of biometric research: accomplishments. Chall Oppor Pattern Recogn Lett 79:80–105
6. Daugman J (2007) Evolving methods in iris recognition. In: IEEE international conference on biometrics: theory, applications, and systems, (BTAS07)
7. Daugman J (2004) How iris recognition works. IEEE Trans Circuits Syst Video Technol 14(1):21–30
8. Hosseini MS, Araabi BN, Soltanian-Zadeh H (2010) Pigment melanin: pattern for iris recognition. IEEE Trans Instrum Meas 59(4):792–804
9. Dong W, Sun Z, Tan T (2009) A design of iris recognition system at a distance. In: Chinese conference on pattern recognition, (CCPR 2009), pp 1–5
10. Proenca H, Filipe S, Santos R, Oliveira J, Alexandre LA (2010) The UBIRIS.v2: a database of visible wavelength iris images captured on-the-move and at-a-distance. IEEE Trans Pattern Anal Mach Intell 32(8):1529–1535
11. Bowyer KW (2012) The results of the NICE. II Iris biometrics competition. Pattern Recogn Lett 33(8):965–969
12. De Marsico M, Nappi M, Riccio D, Wechsler H (2015) Mobile iris challenge evaluation (MICHE)-I, biometric iris dataset and protocols. Pattern Recogn Lett 57:17–23
13. Jinyu Z, Nicolo F, Schmid NA (2010) Cross spectral iris matching based on predictive image mapping. In: Fourth IEEE international conference on biometrics: theory applications and systems (BTAS'10), pp 1–5
14. Nalla PR, Kumar A (2017) Toward more accurate iris recognition using cross-spectral matching. IEEE Trans Image Process 26(1):208–221
15. Du Y, Bourlai T, Dawson J (2016) Automated classification of mislabeled near-infrared left and right iris images using convolutional neural networks. In: 2016 IEEE 8th international conference on biometrics theory, applications and systems, pp 1–6
16. Minaee S, Abdolrashidiy A, Wang Y (2016) An experimental study of deep convolutional features for iris recognition. In: 2016 IEEE signal processing in medicine and biology symposium (SPMB), pp 1–6
17. Raja KB, Raghavendra R, Venkatesh S, Busch C (2017) Multi-patch deep sparse histograms for iris recognition in visible spectrum using collaborative subspace for robust verification. Pattern Recogn Lett 91:27–36
18. Boyce C, Ross A, Monaco M, Hornak L, Xin L (2006) Multispectral iris analysis: a preliminary study. Computer Vision and Pattern Recognition, Workshop
19. Tajbakhsh N, Araabi BN, Soltanianzadeh H (2008) Feature fusion as a practical solution toward noncooperative iris recognition. In: 11th international conference on information fusion, pp 1–7
20. Ngo HT, Ives RW, Matey JR, Dormo J, Rhoads M, Choi D (2009) Design and implementation of a multispectral iris capture system. In: Asilomar conference on signals, systems and computers, pp 380–384
21. Ross A, Pasula R, Hornak L (2009) Exploring multispectral iris recognition beyond 900 nm. In: IEEE 3rd international conference on biometrics: theory, applications, and systems, (BTAS'09), pp 1–8

22. Ives RW, Ngo HT, Winchell SD, Matey JR (2012) Preliminary evaluation of multispectral iris imagery. In: IET conference on image processing (IPR 2012), pp 1–5
23. Sharma A, Verma S, Vatsa M, Singh R (2014) On cross spectral periocular recognition. In: IEEE international conference on image processing (ICIP), pp 5007–5011
24. Alonso-Fernandez F, Mikaelyan A, Bigun J (2015) Comparison and fusion of multiple iris and periocular matchers using near-infrared and visible images. In: 2015 international workshop on biometrics and forensics (IWBF), pp 1–6
25. Crihalmeanu SG, Ross AA (2016) Multispectral ocular biometrics, pp 355–380. Springer International Publishing
26. Burge MJ, Monaco MK (2009) Multispectral iris fusion for enhancement, interoperability, and cross wavelength matching, vol 7334. SPIE, pp 73341D–1–73341D–8
27. Burge M, Monaco M (2013) Multispectral iris fusion and cross-spectrum matching. In: Advances in computer vision and pattern recognition, book section 9. Springer, London, pp 171–181
28. Abdullah MAM, Chambers JA, Woo WL, Dlay SS (2015) Iris biometric: is the near-infrared spectrum always the best? In: 3rd Asian conference on pattern recognition (ACPR2015), pp 816–819
29. Ramaiah NP, Kumar A (2016) Advancing cross-spectral iris recognition research using bi-spectral imaging. In: Advances in intelligent systems and computing, vol 390, book section 1. Springer, pp 1–10
30. Abdullah MAM, Dlay SS, Woo WL, Chambers JA (2016) A novel framework for cross-spectral iris matching. IPSJ Trans Comput Vis Appl 8(9):1–11
31. Al-Nima RR, Dlay SS, Woo WL (2014) A new approach to predicting physical biometrics from behavioural biometrics. In: 16th international conference on image analysis and processing, pp 1–6
32. Abdullah MAM, Dlay SS, Woo WL, Chambers JA (2016) Robust iris segmentation method based on a new active contour force with a noncircular normalization. IEEE Trans Syst Man Cybern Syst PP(99):1–14
33. Kannala J, Rahtu E (2012) BSIF: binarized statistical image features. In: 21st international conference on pattern recognition (ICPR), pp 1363–1366
34. Štruc V, Pavesic N (2009) Gabor-based kernel partial-least-squares discrimination features for face recognition. Informatica 20(1):115–138
35. Štruc V, Pavesic N (2011) Photometric normalization techniques for illumination invariance, pp 279–300. IGI Global
36. Masek L, Kovesi P (2003) MATLAB source code for a biometric identification system based on iris patterns
37. Meredith P, Sarna T (2006) The physical and chemical properties of eumelanin. Pigment Cell Res. 19(6):572–594
38. Raja KB, Raghavendra R, Vemuri VK, Busch C (2015) Smartphone based visible iris recognition using deep sparse filtering. Pattern Recogn Lett 57(C):33–42
39. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987
40. Maltoni D, Maio D, Jain A, Prabhakar S (2003) Multimodal biometric systems, pp 233–255. Springer, New York
41. He M, Horng S-J, Fan P, Run R-S, Chen R-J, Lai J-L, Khan MK, Sentosa KO (2010) Performance evaluation of score level fusion in multimodal biometric systems. Pattern Recogn 43(5):1789–1800
42. Wild P, Radu P, Ferryman J (2015) On fusion for multispectral iris recognition. In: 2015 international conference on biometrics (ICB), pp 31–37
43. Dame A, Marchand E (2012) Second-order optimization of mutual information for real-time image registration. IEEE Trans Image Process 21(9):4190–4203
44. Cheng P, Menq CH (2013) Real-time continuous image registration enabling ultraprecise 2-D motion tracking. IEEE Trans Image Process 22(5):2081–2090

Chapter 6
Facial Surveillance and Recognition in the Passive Infrared Bands

Nnamdi Osia, Thirimachos Bourlai and Lawrence Hornak

Abstract This chapter discusses the use of infrared imaging to perform surveillance and recognition where the face is used for recognizing individuals. In particular, it explores properties of the infrared (IR) band, effects of indoor and outdoor illumination on face recognition (FR) and a framework for both homogeneous and heterogeneous FR systems using multi-spectral sensors. The main benefit of mid-wave infrared and long-wave infrared (MWIR, LWIR) camera sensors is the capability to use FR systems when operating in difficult environmental conditions, such as in low light or complete darkness. This allows for the potential to detect and acquire face images of different subjects without actively illuminating the subject, based on their passively emitted thermal signatures. In this chapter, we demonstrate that by utilizing the "passive" infrared band, facial features can be captured irrespective of illumination (e.g. indoor vs. outdoor). For homogeneous FR systems, we formulate and develop an efficient, semi-automated, direct matching-based FR framework, that is designed to operate efficiently when face data is captured using either visible or IR sensors. Thus, it can be applied in both daytime and nighttime environments. The second framework aims to solve the heterogeneous, cross-spectral FR problem, enabling recognition in the MWIR and LWIR bands based on images of subjects in the visible spectrum.

N. Osia (✉) · T. Bourlai
West Virginia University, PO Box 6201, Morgantown, WV 26506, USA
e-mail: nosia@mix.wvu.edu

T. Bourlai
e-mail: thirimachos.bourlai@mail.wvu.edu

L. Hornak
University of Georgia, 132 Paul D. Coverdell Center, 500 D.W. Brooks Drive,
Athens, GA 30602, USA
e-mail: lahornak@uga.edu

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_6

Introduction

The electromagnetic spectrum spans a wide range of frequencies and corresponding wavelengths. We interact daily with natural sources or manmade systems that emit or utilize these waves, including X-rays; ultraviolet, visible, and infrared (IR) light; as well as millimeter waves, microwaves and radio waves [12]. The longer wavelength infrared bands hold special interest for biometric recognition since all individuals possess infrared emission signatures corresponding to the spatial distribution of their body temperatures that is emitted passively, without external illumination. For the purposes of this chapter and in the context of facial recognition utility, we refer to these as the passive IR bands.

Differences in appearance arise between images sensed in the visible and the active IR band, primarily due to the properties of the object being imaged. The higher frequency, smaller wavelength part of the IR spectrum is composed of the Near IR band (0.7–0.9 μm) and the Short-Wave IR band (0.9–2.5 μm). From a recognition perspective, subjects must be illuminated actively by sources with output at these IR wavelengths and a reflection return from the illumination received at the camera for imaging. At lower frequencies and longer wavelengths, the IR spectrum consists of the Mid-Wave IR (MWIR) (3–5 μm), and Long-Wave IR (LWIR) (7–14 μm) bands. In these IR spectral bands, radiation is passively emitted from the target based on its temperature, in this particular case the subject's face, and detected by arrays of sensors during image acquisition. IR sensors in the MW and LW bands are beneficial in challenging conditions, and provide the added benefit of not requiring added illumination which may otherwise be detected. The combination of IR imaging sensors at different bands utilizing active and passive illumination properties may hold the potential to help improve FR accuracy where illumination may be an uncertainty.

This chapter can be further divided into three separate sections. Section "Literature Review" details related work and face matching approaches in the passive IR band. Section "Indoor and Outdoor Illumination on Face Recognition" includes indoor and outdoor illuminated face recognition analysis. In section "Homogeneous Face Recognition", we formulate a face matching pipeline for same spectrum matching. Section "Heterogeneous Face Recognition" looks at the more challenging heterogeneous FR scenario where we are utilizing image synthesis for cross-spectrum matching. We summarize the chapter in section "Conclusion".

Literature Review

Pan et al. [13], is one of the earliest works on hyperspectral FR. Pan et al. conducted their work on over 30 bands in the NIR region. The authors utilized local spectral properties of human tissue that is robust to face orientation and expression, allowing hyperspectral discriminants to be used for recognition over a large range of poses and expressions. In [10] the authors reported that FR in the LWIR band achieves a rank-1 accuracy of 97.3% when using local binary patterns (LBP), while no cropping or geometrical normalization step is required. In Socolinsky et al. [17], the authors used two standard FR algorithms to show that, under variable illumination conditions, the usage of LWIR face images yield a higher recognition performance than visible ones. However, the drawback of the approaches is that LWIR and visible images were divided into multiple training and testing sets, resulting in an increase of the FR system design time. In addition, [17] was performed using co-registered images that were captured simultaneously by both visible and LWIR cameras—this is not usually possible in operational environments. In other IR-based FR approaches, such as Trujillo et al. [20], the authors proposed an unsupervised local and global feature extraction paradigm to classify different facial expressions. In Chen et al. [5] the authors combined visible and thermal-based images and compared them using Principle Component Analysis. However, neither Trujillo et al. [20] nor *Chen's* work [5] focused on the MWIR band. Over the last couple of years, MWIR and LWIR sensing technology has grown in terms of resolution, pixel size and advances in methodological approaches (Figs. 6.1, 6.2 and 6.3).

Recent advances in appearance based IR FR has closely reflected research in visible spectrum based recognition. Progress in comparison with some aforementioned earlier works is rooted mainly in the use of more sophisticated statistical techniques. For example, Elguebaly et al. [6] recently described a method based on a generalized Gaussian mixture model, where parameters are learned from a training image set using a Bayesian approach. Although substantially more complex, this approach did

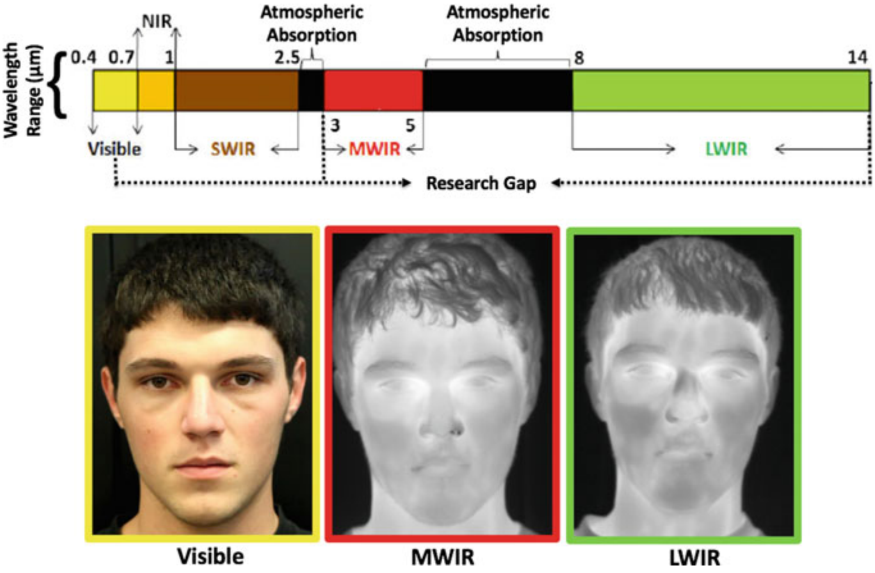


Fig. 6.1 Sample face images of a randomly selected subject are also illustrated that correspond to the three spectral bands of interest, i.e. **a** Visible (yellow), **b** MWIR (red), **c** LWIR (green)

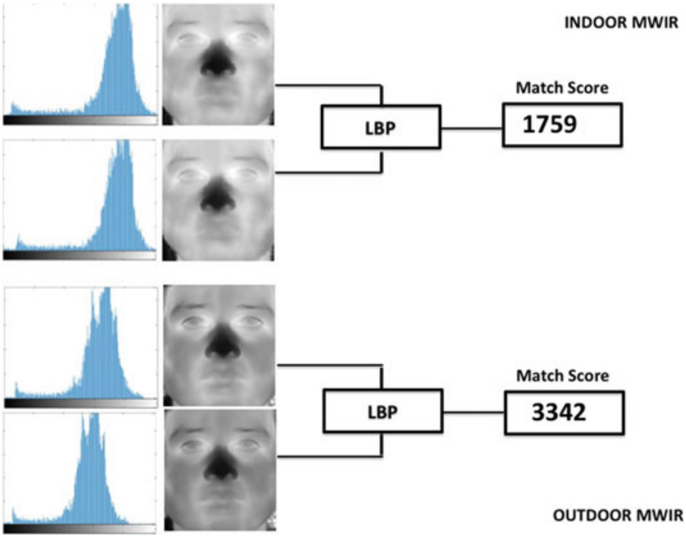


Fig. 6.2 Histogram and normalized face images of different samples for the same subject. LBP or difference scores between the two normalized face samples using (top) Indoor and (bottom) Outdoor Environment in the MWIR band

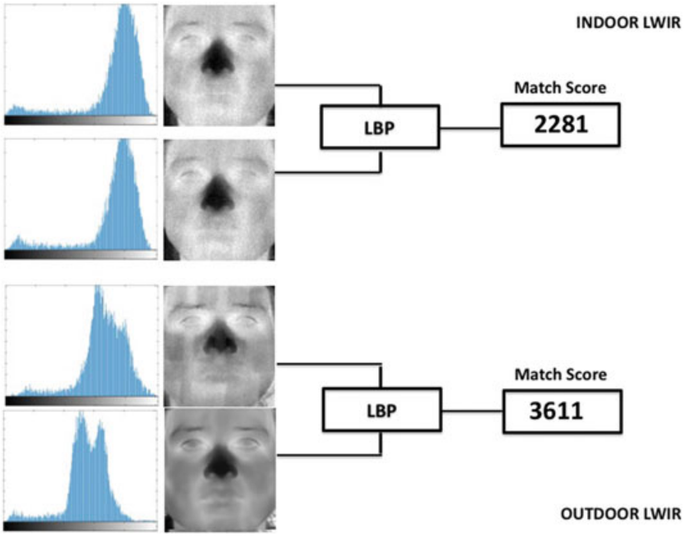


Fig. 6.3 Histogram and normalized face images of different samples for the same subject. LBP or difference scores between the two normalized face samples using (top) indoor and (bottom) outdoor Environment in the LWIR band

not demonstrate a statistically significant improvement in recognition for database used, achieving rank-1 rate of approximately 95%. The wavelet transform has been studied extensively as a means of representing a wide range of 1D and 2D signals, including face appearance in the visible spectrum, because of its ability to capture both frequency and spatial information. Srivastava et al. [18] were the first to investigate use of wavelet transforms based on a bank of Gabor filters for extracting robust features from face appearance images in the IR spectrum. The marginal density functions of the filtered features are then modelled using *Bessel K forms* which are matched using the simple L_2-norm. Srivastava et al. reported a remarkable fit between the observed and the estimated marginals across a large set of filtered images. The curvelet transform is an extension of the wavelet transform, in which the degree of orientational localization is dependent on the scale of the curvelet [8]. The curvelet transform facilitates a sparser representation than wavelet transforms with effective spatial and directional localization of edge-like structures in natural images. Xie et al. [23] described the first IR based FR system which utilizes the curvelet transform for feature extraction. The method utilized a simple nearest neighbor classifier which demonstrated a slight advantage (of approximately 1–2 %) over simple linear discriminant based approaches, but with a significant improvement in computational and storage demands. Xie et al. [21] and Wu et al. [22] proposed to exploit temperature differential between vascular and non-vascular tissues, extracting invariant features in IR imagery. Wu et al. formulated the model governing blood perfusion, which is based on differential equations by using a series of assumptions of relative temperatures of the body's deep and superficial tissues, and the ambient temperature. The model is then used to compute a "blood perfusion image" from the original segmented thermogram of a face. Finally, blood perfusion images are matched using a standard linear discriminant and a network of radial basis functions. These works are summarized in Table 6.1.

Indoor and Outdoor Illumination on Face Recognition

It has been qualitatively observed that IR imagery of human faces is invariant to changes in indoor and outdoor illumination, although there hasn't been much quantitative analysis to confirm this in open literature. In the SWIR and NWIR bands, artificial illumination sources can be used to increase contrast when imaging by using reflected energy from faces being imaged. However, an advantage of the passive IR band is that faces imaged within the band are good sources of IR energy, therefore no additional illumination is necessary. If this holds true, there should not be an affect on FR when there is varying illumination on the face being imaged.

In an attempt to determine the effects of varying illumination on FR in the IR band, we collect data samples both indoors and outdoors in the MWIR and LWIR spectrums respectively for one subject. After the face images are captured, they are normalized using geometric normalization techniques. We look at the LBP distance between two samples of the same subject for four different scenarios, as well as

Table 6.1 Literature review matrix for passive infrared band

Author(s)	Spectrum(s)	Approach/features	Pros	Cons
Pan et al. [13]	LWIR	Hyperspectral discriminants	Robust to pose and expression	Hyperspectral cameras necessary
Mendez et al. [10]	LWIR	Local binary patterns (LBP)	Rank-1 accuracy of 97.3%	No cropping or geometric normalization
Socolinsky et al. [17]	Visible and LWIR	Eigen-faces and ARENA (appearance based/l-nearest neighbor)	Co-registered visible and LWIR images captured simultaneously	LWIR and visible divided into multiple training and testing sets
Trujillo et al. [20]	LWIR	Local and global feature extraction paradigm	Unsupervised (no learning necessary)	Facial expressions were classified, no FR
Chen et al. [5]	Visible and LWIR	Principle component analysis (PCA)	Visible and thermal-based images fused	FR accuracy is lower in time-lapse (different session scenarios)
Buddharaju et al. [3]	MWIR	Physiological features	Robust and invariant to pose	No cropping or geometric normalization
Elguebaly and Bouguila [6]	Visible and LWIR	Gaussian mixture model	Unsupervised algorithm with rank-1 accuracy of 95%	Substantially more complex
Srivastana and Liu [18]	Visible and LWIR	Wavelet transform based on Gabor filters	Effective with low-resolution images	Bessel parameters significantly reduce representation
Xie et al. [23]	LWIR	Curvelet transform	Improved computational and storage demands	Better performance with radiant energy instead of thermal
Wu et al. [22]	LWIR	Blood perfusion	Less sensitive to ambient temperature	Time and storage efficiency

the histogram for each subject sample as a quantitative measurement. Our results indicate that varying illumination does not have a significant impact on FR for a single subject. When operating in an indoor environment, we found a match between two subjects to be stronger (smaller LBP score or distance), when compared to that of an outdoor environment, irrespective of spectrum operation.

Homogeneous Face Recognition

For homogeneous FR systems, we formulate and develop an efficient, semi-automated, direct matching-based FR framework, that is designed to operate efficiently when face data is captured using sensors in the visible or passive IR band. Thus, it can be applied in both daytime and nighttime environments. First, input face images are geometrically normalized using our pre-processing pipeline prior to feature-extraction. Then, face-based features including wrinkles, veins, as well as edges of facial characteristics, are detected and extracted for each operational band (visible, MWIR, and LWIR). Finally, global and local face-based matchers are suggested for use in matching the detected features.

Face Image Feature Extraction and Matching Methodology

With regards to the semi-automated pre-processing approach, human eyes and eye centers are detected. Then, eye coordinates are used to geometrically normalize faces. Also, the inter-ocular distance is fixed by setting the dimensions of the input face images at a specific spatial resolution and the eye centers of these images at predetermined (x, y) locations. Prior to the feature extraction stage, the parameters of the image diffusion and face segmentation algorithms are enhanced. Anisotropic image diffusion is also used to smooth each face image, prior to applying top-hat segmentation.

In our experiments, we use the segmented features to demonstrate that our face recognition system performance improves when fusing global and local-based matchers. The global matcher uses the feature segmented image for matching, i.e. this matcher analyzes the morphology of the face through the convolution of face images, producing a match ratio [11]. Although the global matching algorithm is based on overlap of neighborhood pixels, the accuracy of the algorithm increases when unique segments are matched. The local matcher requires fiducial points that we obtain using the extracted feature segment, and matches all points to one another (point to point). These fiducial points are minutiae (level 2 features) extracted from physiologically-based (when using subcutaneous facial characteristics) and geometrically-based face features (e.g. eye edges and eyelashes), which are unique for each individual (see Fig. 6.5). The use of minutia points helps facilitate a metric for measuring similarity using our local matcher and helps improve accuracy through precise selection of features. In the matching and decision making step for both matchers, after our feature segmentation scheme is applied to each image, input images (probes) are matched with a stored template (gallery).

Pre-processing

Overview of the methodology used to perform pre-processing, passive IR-based feature extraction, and matching can be seen in Fig. 6.4.

- *Inter-ocular Normalization*: It standardizes all face images so head sizes are relatively similar. Once the eyes are manually detected on the passive IR images, they are used to normalize all images so that the inter-ocular distance is fixed to 60 pixels. This is accomplished by resizing the image acquired using a ratio computed from the desired inter-ocular distance (60 px) and the actual inter-ocular distance, i.e. the one computed when using the raw image.
- *Geometric Normalization*: A geometric normalization scheme is applied to images acquired after inter-ocular normalization. The normalization scheme compensates for slight perturbations in the frontal pose, and consists of eye detection and affine transformation. After the eye centers are found using manual eye detection, the canonical faces are automatically constructed by applying an affine transformation. Finally, all faces are canonicalized to the same dimension of 320 × 256.

  The proposed methodology is composed of a feature extraction and matching process. The feature extraction process has two steps. First, anisotropic diffusion is used to reduce noise while preserving vital image content. Next, top hat segmentation is carried out in order to segment and extract face-based features. The features extracted include: (a) veins, (b) edges, (c) wrinkles, and (d) face perimeter outlines (see Fig. 6.5). After features are extracted, different matchers (e.g. global and local) are utilized, before finally, they are fused together at the score level in an effort to achieve increased rank-1 identification performance. The aforementioned methodological steps are described below.

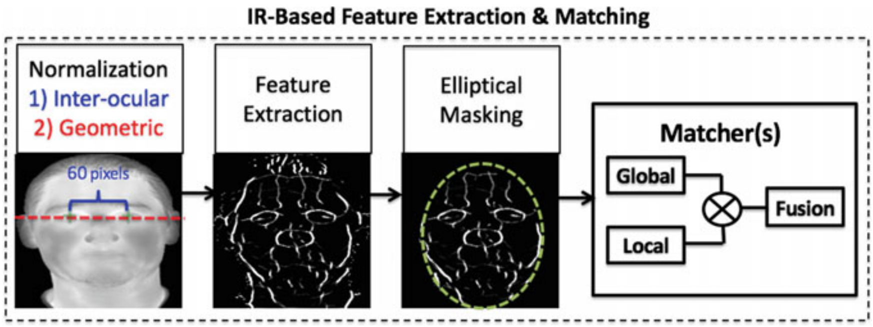


Fig. 6.4 Overview of the methodology used to perform pre-processing, passive IR-based feature extraction, and matching

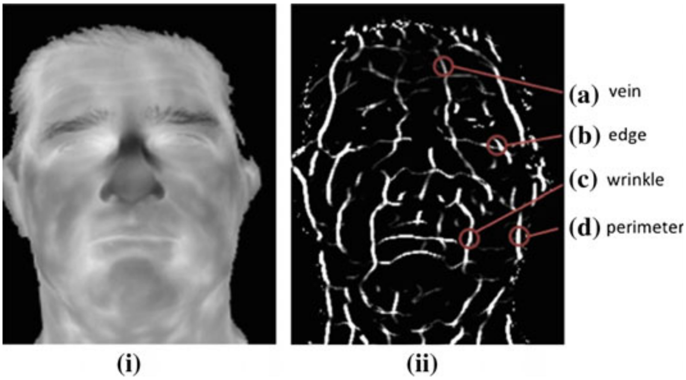


Fig. 6.5 (i) Geometrically normalized face (before elliptical masking); (ii) Diffused and top hat segmented face (before elliptical masking): **a** veins, **b** edge, **c** wrinkle and **d** part of the face perimeter

Anisotropic Diffusion and Top Hat Segmentation

Perona-Malik's anisotropic diffusion is used to process the face images and remove background noise [14]. This is important because noise is reduced without the removal of significant image content, such as edges and lines. The mathematical representation for this process is described as follows:

$$\frac{\partial I(\bar{x},t)}{\partial} = \nabla(c(\bar{x},t)\nabla I(\bar{x},t)) \tag{6.1}$$

where $I_{N,t} = I_t(x, y + 1) - I_t(x, y)$. The diffusion operator inherently behaves differently, depending on which spectrum we are operating in. For images in the visible spectrum, diffusion is used in edge detection between lines of dissimilar contrast. The face-based features (wrinkles, veins, edges, and perimeters) are segmented by the use of image morphology. For the passive infrared band, heat diffusion generally produces weak sigmoid edges in the thermal band during heat conduction, which in turn creates smooth temperature gradients at the intersecting boundary of multiple objects with dissimilar temperatures in contact. In order to segment these features, morphological top-hat filtering is employed.

$$I_{open} = (I \ominus S) \bigoplus S, \tag{6.2}$$

$$I_{top} = I - I_{open} \tag{6.3}$$

I, I_{open}, and I_{top} are the original, opened, and white top hat segmented images, respectively. S is the structuring element, and $\ominus$ and $\oplus$ are the morphological operations for erosion and dilation respectively. Parameters for Perona-Malik anisotropic diffusion process along with top hat segmentation, were empirically optimized to ensure the resultant images (see Fig. 6.5 (ii)) did not contain noise, i.e. outlier edges that do not represent clear face-based physiological and geometrical features. Pixel normalization was the only pre-processing done on the images during this experiment, unlike the original work, which dilated, skeletonized, and bridged MWIR images [3].

Elliptical Masking

It is imperative that the elliptical mask is applied after the feature extraction stage to ensure that spurious, artificial feature points are not created by the mask's presence during top hat segmentation. In practice, image masking ensures no clothing is present during feature detection. When the elliptical mask segments the face, the part of the original image that is masked is set to a black background. Note also that the dimensions of the ellipse used for all images were fixed.

Matching

Many FR matching methods can be categorized as being global or local, depending on whether features are extracted from the entire face or from a set of local regions. Global features are usually represented by vectors of fixed in length, which are compared during the matching process in a time efficient manner. On the contrary, local feature-based approaches first detect a set of fiducial points, using the surrounding pixel values. The number of matched fiducial points between an input and gallery template is used to calculate the match score. Since the number of fiducial points may vary depending on the subject, two sets of points from two different subjects cannot be compared directly. As a result, the matching scheme has to compare each fiducial point from one template against all the other fiducial points in another template, increasing time for matching. Global features are more susceptible to variations in the face when all pixel values are encoded into a feature vector, especially with respect to geometric transformations.

- *Global Matcher (Segmentation Matching)*: It is usually a difficult task to extract on-face information using only simple techniques. Our feature segmentation step extracts lines, edges, and wrinkles, which are unique to each subject, such as in the case of contours. In two face images of the same subject, similar features may still be found in contours. However, there are generally remarkable differences in not only the shape of these contours but in the size as well, across subjects. Hence, identification may be carried out using matching of the segmented features. The segmented face features are compared using template matching.
  To find the maximum similarity between a gallery and probe image, the two feature segmented images are matched. The probe image is slid pixel by pixel across the gallery image in a top to bottom and left to right fashion. If $f(i,j)$ and $g(k,l)$ are the pixels at position (i,j) and (k,l) of probe and gallery feature segments

respectively, then α, β measure the horizontal and vertical displacement between $f(i,j)$ and $g(k, l)$ during the traversal process. To account for the small differences that exist in the segmented features from different samples of the same subject, a 5×5 window around a white pixel is used for template matching. If for a white pixel in the probe segmented image, there is another white pixel in the 5×5 neighborhood of the corresponding position on the gallery segmented image, then the pixels are said to be matched. If $\hat{\alpha}, \hat{\beta}$ are the horizontal and vertical displacement respectively, which give the best matching result, then the maximum similarity $h_{i,j}(\hat{\alpha}, \hat{\beta})$ defined by (6.4) between corresponding feature segments can be obtained [11].

$$h_{i,j}(\hat{\alpha}, \hat{\beta}) \overset{\triangle}{=} \max_{\alpha,\beta} \sum_{i,j} h_{i,j}(\alpha, \beta) \tag{6.4}$$

$$h_{i,j}(\alpha, \beta) = \phi[\sum_{x=-2}^{2} \sum_{y=-2}^{2} (f_{i,j} \cdot g_{k+x,l+y})] \tag{6.5}$$

$k = i \pm \alpha, l = j \pm \beta,$
$\alpha = 0, 1, 2, \ldots, 320, \beta = 0, 1, 2, \ldots, 256$

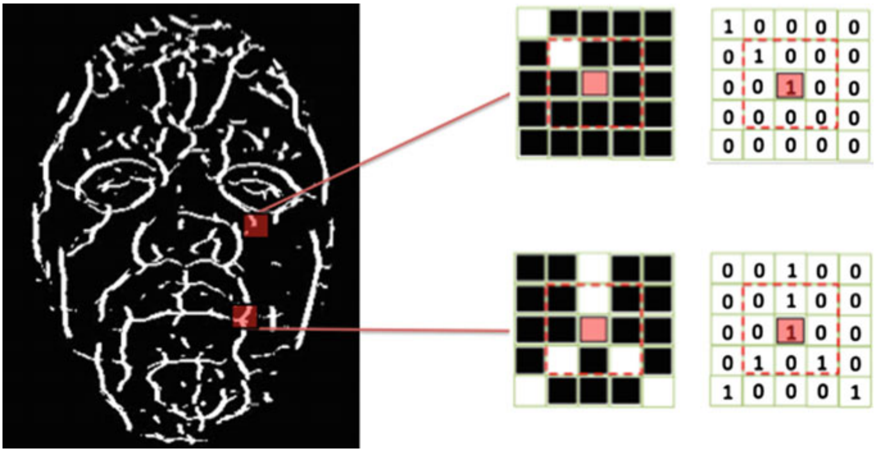
$$\phi[x] = \begin{cases} 1, & \text{for } x \geq 1 \\ 0, & \text{for } x = 0 \end{cases} \tag{6.6}$$

For images of the same subject, after template matching, long overlapping segments (in the sense of (6.5)) are obtained. On the other hand, for different subjects, even if a high match score is obtained from (6.4), the existence of abundant segments overlapped by chance is expected. Therefore, if these short segments can be eliminated effectively, a stable performance of the discrimination method can be achieved. The final matching score $H(f, g)$ can be calculated using (6.7).

$$H_{f,g} = \frac{2}{F+G} \cdot \sum_{i,j} h_{i,j}(\hat{\alpha}, \hat{\beta}) \tag{6.7}$$

F and G denote the number of pixels in the feature segmented lines of the probe and gallery images respectively. The fragment removal threshold θ_i was optimized experimentally, and held constant for each image. Prior to matching, all images are filtered by removing pixel values below θ_i.

- *Local Matcher (Fiducial Point Matching)*: The proposed method, a fingerprint based minutia point recognition system is used to detect features in the face [24]. Beforehand, the pixel intensities of the images are thresholded so that they are binary. For our normalization, the mean pixel intensity of the image is set as the threshold for passive infrared images. While traversing the image, if the current pixel intensity is smaller than the threshold, it receives a binary value of zero. If the current pixel intensity is larger than the threshold, it receives a binary value



1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	0	0
0	0	0	0	0

0	0	1	0	0
0	0	1	0	0
0	0	1	0	0
0	1	0	1	0
1	0	0	0	1

Fig. 6.6 Sample feature segmented subject face with **a** block representation of end point pixels, **b** binary representation of endpoint, **c** block representation of branch minutia point pixels and **d** binary representation of branch minutia point. Red box denotes marked minutia point

of one. During the traversal process, if the 3 × 3 neighborhood around a center white pixel had exactly three white pixels, it was labeled a branch point. If the center white pixel with a 3 × 3 neighborhood around it contained only 1 white pixel neighbor, then it was labeled an end point (see Fig. 6.6). Depending on the normalization technique, large clusters of both branch points and end points may be found.

A point alignment-based Random Sample Consensus (RANSAC) matching algorithm with the ability of finding the correspondences between a stored set of gallery points and input set of probe points is used. The set of input points are first aligned with the gallery points and then a match score is computed, based on corresponding points. Let G and P be the gallery and probe points we are trying to match. Because the correspondence points are computed based on local information of each subject, G and P may not have the same number of correspondence points: let m and n be the number of points for G (6.8) and P (6.9).

$$G = \{c_1, c_2, \ldots c_m\}, \text{ with } c_i = \{x_i, y_i\}, i = 1 \ldots m \tag{6.8}$$

$$P = \{c_1', c_2', \ldots c_n'\}, \text{ with } c_j' = \{x_j', y_j'\}, j = 1 \ldots n \tag{6.9}$$

Two correspondence points c_i and c_i' are considered matching if close in position. This can be written according to (6.10), using the spatial distance sd. Two minutia points from each set are considered matching if their spatial distance (sd) is within a threshold (R_θ).

$$sd(c_i, c_j') = \sqrt{(x_i - x_j')^2 + (y_i - y_j')^2} < R_\theta \tag{6.10}$$

The distance between each possible point in the gallery set and point in the probe set is computed and sorted in increasing order. We then count the number of matches below a threshold, which is optimized for the fixed dimension of our face images. The highest counted number of correspondences; C_{best}, for each set of points is stored and used to compute the match score, M_S. We let:

$$M_S = \frac{(C_{best})^2}{tot\{G,P\}}. \tag{6.11}$$

Heterogeneous Face Recognition

The formulation of the cross-spectral matching problem that we want to solve is as follows. Given a IR image I_p as input, we estimate the target VIS image V_p with the help of a training set of IR images I_q and the corresponding VIS images V_q. We represent each IR or VIS image as a set of small image patches that overlap. I_p and V_p have the same number of patches, and each IR image in I_q and the corresponding VIS image in V_q also have the same number of patches.

Utilizing the IR patches, linear models are used to predict their corresponding VIS patches. An input IR patch is represented as $y \in \Re^D$, D is its pixel number. Then from each training IR-VIS image pair, we extract patches of the same size at the same position as y. These IR and VIS patches are denoted as Y and X respectively. To predict y's VIS counterpart x, relations learned from X and Y are used. Following the assumption of image manifolds [1, 4, 7], we take Y and X as samples drawn from two manifolds. Therefore, the neighborhood of y in Y can be seen as lying on a linear subspace. A search is conducted for the K nearest neighbors in Y for y, Y_N. Their corresponding VIS patches are denoted as X_N. Canonical Correlation Analysis (CCA) is used to model the linear relations between the linear subspaces spanned by Y_N and X_N. CCA discovers one set of axes for each dataset, along which these two sets of data co-vary most. From the viewpoint of learning, CCA finds the most linear predictable components for the two sets.

In the proposed approach, patch size and degree of overlap between adjacent patches are taken into consideration. Ideally, each patch generated for the VIS image V_p should be related appropriately to the corresponding patch in the IR image I_p and also preserve some inter-patch relationships with adjacent patches in V_p. In Fig. 6.7, we draw a flowchart for the proposed solution to our cross-spectral matching problem.

Image Synthesis

The formulated image synthesis methodology is combination of manifold learning and non-linear dimensionality reduction. We utilize the leave one out method during

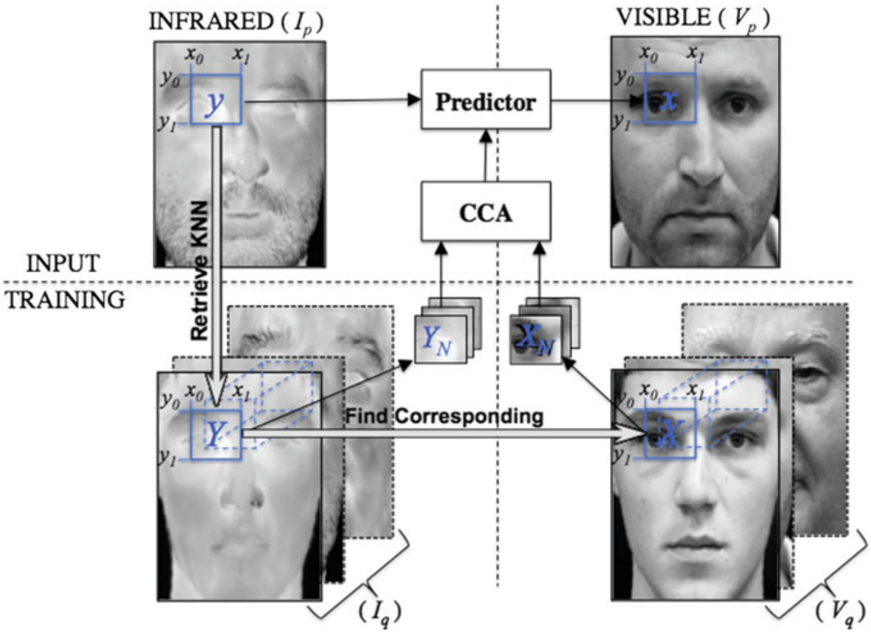


Fig. 6.7 Flow chart of proposed image synthesis

synthesis, the sample left out of the training set is used for conversion from one spectrum to another. Through the image synthesis algorithm, we are able to convert the datasets described and create their synthesized versions. After the synthesized data is created, it is later used for identity authentication.

Canonical Correlation Analysis

Through the use of two random variables with zero-mean **x**, a $p \times l$ vector, and **y**, a $q \times l$ vector, CCA finds the *1*st pair of directions $\mathbf{w}_1$ and $\mathbf{v}_1$ that results in the greatest correlation between the projections $\mathrm{x} = \mathbf{w}_1{}^{\mathrm{T}}\mathbf{x}$ and $\mathrm{y} = \mathbf{v}_1{}^{\mathrm{T}}\mathbf{y}$, max $\rho(\mathbf{w}_1{}^{\mathrm{T}}\mathbf{x},\ \mathbf{v}_1{}^{\mathrm{T}}\mathbf{y})$, *s.t.* $Var((\mathbf{w}_1{}^{\mathrm{T}}\mathbf{x} = 1)$ and $Var(\mathbf{v}_1{}^{\mathrm{T}}\mathbf{y} = 1)$, where the correlation coefficient is ρ, the variables x and y are known as the first canonical variates, and the $\mathbf{w}_1$and$\mathbf{v}_1$ represents the initial correlation direction vector. CCA finds *k*th pair of directions $\mathbf{w}_k$ and $\mathbf{v}_k$ which satisfies: (1) $\mathbf{w}_k{}^{\mathrm{T}}\mathbf{x}$ and $\mathbf{v}_k{}^{\mathrm{T}}\mathbf{y}$ are not correlated to the previous *k-1* canonical variates; (2) the correlation between $\mathbf{w}_k{}^{\mathrm{T}}\mathbf{x}$ and $\mathbf{v}_k{}^{\mathrm{T}}\mathbf{y}$ is optimized under the constraints $Var((\mathbf{w}_1{}^{\mathrm{T}}\mathbf{x} = 1)$ and $Var(\mathbf{v}_1{}^{\mathrm{T}}\mathbf{y} = 1)$. Then $\mathbf{w}_k{}^{\mathrm{T}}\mathbf{x}$ and $\mathbf{v}_k{}^{\mathrm{T}}\mathbf{y}$ are called the k^{th} canonical variates, and $\mathbf{w}_k$ and $\mathbf{v}_k$ are the k^{th} correlation direction vector, $\mathrm{k} \leq \min(p,\ q)$. The solution for the correlation of coefficients and directions is not different from the generalized eigenvalue problem seen here,

$$(\Sigma_{\mathrm{xy}}{\Sigma_{\mathrm{yy}}}^{-1}{\Sigma_{\mathrm{xy}}}^{T} - \rho^2\Sigma_{\mathrm{xx}})\mathbf{w} = 0, \tag{6.12}$$

$$({\Sigma_{\mathrm{xy}}}^{T}{\Sigma_{\mathrm{xx}}}^{-1}\Sigma_{\mathrm{xy}} - \rho^2\Sigma_{\mathrm{yy}})\mathbf{v} = 0\ , \tag{6.13}$$

where Σ_{xx} and Σ_{yy} are the self-correlation while the Σ_{xy} and Σ_{yx} are the co-correlation matrices respectively. Through CCA, the correlation of the two data sets are prioritized, unlike PCA, which is designed to minimize the reconstruction error. Generally speaking, a few projections (canonical variates) are not adequate to recover the original data well enough, so there is no guarantee that the directions discovered through CCA cover the main variance of the paired data. In addition to the recovery problem, the overfitting problem should be accounted and taken care of as well. If a small amount of noise is present in the data, CCA is so sensitive it might produce a good result to maximize the correlations between the extracted features, but the features may likely model the noise rather than the relevant information in the input data. In this work we use a method called regularized CCA [9]. This approach has proven to overcome the overfitting problem by adding a multiple of the identity matrix $\lambda\mathbf{I}$ to the co-variance matrix Σ_{xx} and Σ_{yy}.

Feature Extraction Using CCA

Local features are extracted, instead of features that are holistic, because the latter features seem to fail capturing localized characteristics and facial traits. The datasets used in training CCA consists of paired VIS and IR images. The images are divided into patches that overlap by the same amount at each position, where there exists a set of patch pairs for CCA learning. CCA locates directional pairs $\mathbf{W}^{(i)} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k]$ and $\mathbf{V}^{(i)} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$ for VIS and IR patches respectively, where the superscript (i) represents the index of the patch (or the location of the patch within the face image). Each column of $\mathbf{W}$ or $\mathbf{V}$ is a directionary vector, which is unitary, but between different columns it is not orthogonal. For example, if we take a VIS patch $\mathbf{p}$ (which can be vectorized as a column) at position i, we are able to extract the CCA feature of the patch $\mathbf{p}$, using $\mathbf{f} = \mathbf{W}^{(i)T}\mathbf{p}$, where $\mathbf{f}$ is the feature vector belonging to the patch. For each patch and each position at each patch, we are able to acquire CCA projections using our preprocessed training database face images. Projection onto the proper directions is used to extract features, then at each patch location i we get the VIS ${\mathbf{O}_v}^i = \{{\mathbf{f}_{v,j}}^i\}$ and IR training sets ${\mathbf{O}_{ir}}^i = \{{\mathbf{f}_{ir,j}}^i\}$ respectively.

Reconstruction Using Training Patches

In our reconstruction phase that occurs during testing, we use explicitly learned LLE weights in conjunction with our training data to reconstruct the patch and preserve the global manifold structure. Reconstructing the original patch $\mathbf{p}$ through the vectorized feature $\mathbf{f}$ is an arduous task. We are unable to recover the patch by $\mathbf{p} = \mathbf{Wf}$ as we do in PCA because $\mathbf{W}$ is not orthogonal. However, the original patch can be obtained by solving the least squares problem below,

$$\mathbf{p} = \mathbf{arg}_p\mathbf{min}||\mathbf{W}^T\mathbf{p} - \mathbf{f}||_2{}^2, \tag{6.14}$$

or to add an energy constraint,

$$\mathbf{p} = \mathbf{arg}_p\mathbf{min}||\mathbf{W}^T\mathbf{p} - \mathbf{f}||_2{}^2 + ||\mathbf{p}||_2{}^2. \tag{6.15}$$

The least squares problem can be solved effectively using the scaled conjugate gradient method. In order for the above reconstruction method to be feasible, the feature vector $\mathbf{f}$ has to contain enough information about the original patch. The original patch can be recovered using LLE [16] when fewer features, represented as canonical variates, can be extracted. The assumption that localized geometries pertaining to the manifold of the feature space and that of the patch space are similar, is taken into consideration (see [4]). The patch from the image to be converted and its corresponding features have similar reconstruction coefficients. If $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_k$ are the patches whose features $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k$ are $\mathbf{f}'\mathbf{s}$ k nearest neighbors, and $\mathbf{f}$ is able to be recovered using neighboring features with $\mathbf{f} = \mathbf{Fw}$, where $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k]$, $\mathbf{w} = [\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_k]^{\mathbf{T}}$, we can reconstruct the original patch using $\mathbf{p} = \mathbf{Pw}$, where $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_k]$. Using a probe IR image, we partition it into small patches, and obtain the feature vector $\mathbf{f}_{ir}$ of every patch. When we infer the corresponding VIS feature vector $\mathbf{f}_v$, the VIS patch can be obtained using $\mathbf{p} = \mathbf{Pw}$ for reconstruction and then the patches will be combined into a VIS facial image. A sample illustration of the reconstruction process can be seen in Fig. 6.8 for K = 5 nearest neighbors.

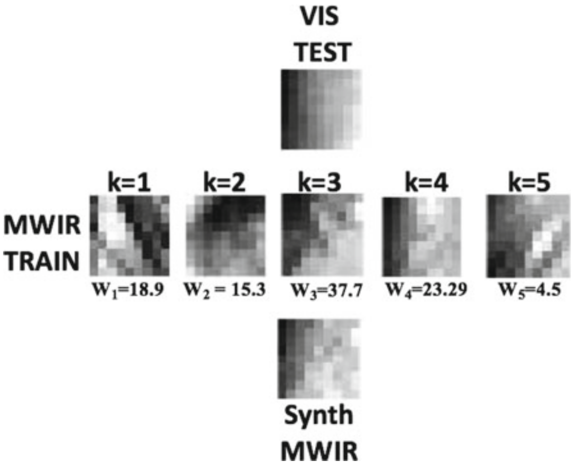


Fig. 6.8 Sample illustration of input VIS patch, and corresponding training MWIR patches, for k = 5 nearest neighbor. The reconstructed and synthesized MWIR image using training patches and locally linear embedded weights

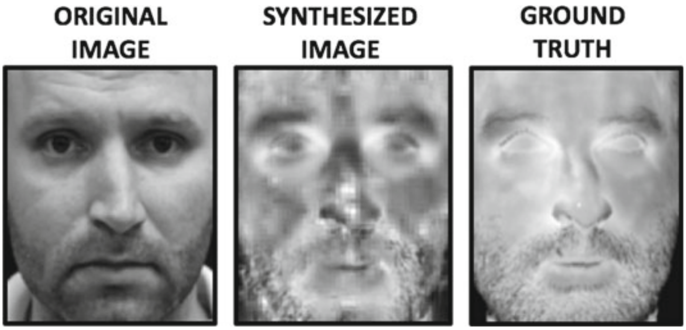


Fig. 6.9 Example original, synthesized, and ground truth images from a subject. The subject's face image is converted from VIS to MWIR and normalized using our proposed pre-processing technique

Methodological Steps

The salient stages of the proposed method are described below (Fig. 6.9):

1. ***Pre-processing***: Our proposed approach is patch-based, therefore it is important that the correct corresponding patches overlap as precisely as possible in both spectra. During pre-processing, a standard interocular distance is set and the eye locations are centered and aligned onto a single horizontal plane and resized to fit the desired distance. Each face image was geometrically normalized based on the manually found locations to have an interocular distance of 60 pixels with a resolution of 111 × 121 pixels. There is no elliptical mask applied in our approach, in contrast to the CSU normalization software.
2. ***Image Synthesis***: The methodology discussed in this section is used. For training during synthesis, the leave one out method is proposed, where the sample left out of the training set is used for conversion from one spectrum to another.
3. ***Face Recognition Matcher***: A number of matchers, both commercial and academic, can be used to evaluate the image synthesis approach. The authors suggest the utilization of a variation of the *Local Binary Patterns* (LBP) method [19] for FR [2]. The LBP operator is an efficient, nonparametric, and unifying approach to traditional divergent models for analyzing texture that are statistical and structural based. A binary code is produced by thresholding the value of the center pixel with its value, for each pixel in an image [15].

Conclusion

In this chapter, we explore characteristics of the infrared band in the context of their utility for Facial Recognition (FR), and determine the effects of indoor and outdoor illumination on FR in the MWIR and LWIR bands. Our indoor versus outdoor captured face experimental results indicate that these environments for acquisition does not have a significant impact on FR in these IR bands for a single subject. For an indoor environment, we found a match between two subjects to be stronger (smaller LBP score or distance), when compared to that of an outdoor environment, irrespective of the spectrum of operation. We also propose a framework, for both homogeneous and heterogeneous FR systems using multi-spectral sensors. For homogeneous FR systems, we formulate and develop an efficient, semi-automated, direct matching-based FR framework, that is designed to operate efficiently when face data is captured using either visible, MWIR or LWIR sensors. Thus, it can be applied in both daytime and nighttime environments. The second framework aims to solve the heterogeneous, cross-spectral FR problem, enabling recognition in the MWIR and LWIR bands based on images of subjects in the visible spectrum.

References

1. Beymer D, Poggio T (2006) Image representation for visual learning. Science
2. Bourlai T, Kalka N, Ross A, Cukic B, Hornak L (2010) Cross-spectral face verification in short infrared band. In: Proceedings of IEEE, International conference on pattern recognition (ICPR), Istanbul, pp 1343–1347
3. Buddharaju P, Pavlidis P, Tsiamyrtzis P, Bazakos M (2007) Physiology-based face recognition in the thermal infrared spectrum. IEEE Trans Pattern Anal Mach Intell 29(4):613–626
4. Chang H, Yeung DY, Xiong Y (2004) Super-resolution through neighbor embedding. In: CVPR
5. Chen X, Flynn P, Bowyer K (2003) PCA-based face recognition in infrared imagery: baseline and comparative studies. In: Proceedings of IEEE international workshop on analytics and modeling of faces and gestures (AMFG). IEEE, pp 127–134
6. Elguebaly T, Bouguila N (2011) A Bayesian method for infrared face recognition. Mach Vis Beyond Visible Spectr 1:123–138
7. Fan W, Yeung DY (2004) Image hallucination using neighbor embedding over visual primitive manifolds. In: CVPR
8. Mandal T, Majumdar A, Wu Q (2007) Face recognition by curvelet based feature extraction. In: ICIAR, pp 806–817
9. Melzer T, Reiter M, Bischof H (2003) Appearance model based on kernel canonical correlation analysis. Pattern Recogn 36:1961–1971
10. Mendez H, Martin C, Kittler J, Plasencia Y, Reyes E (2009) Face recognition with lwir imagery using local binary patterns. In: Proceedings of international conference on advances in biometrics (ICB). Springer, Berlin, pp 327–336
11. Nakamura O, Mathur S, Minami T (1991) Identification of human faces based on isodensity maps. IEEE Proc Pattern Recogn 24(3):263–272
12. NASA (2013) Electromagnetic spectrum. Imagine the universe. http://imagine.gsfc.nasa.gov/science/toolbox/emspectrum1.html. Accessed 19 May 2015
13. Pan Z, Healey G, Prasad M, Tromberg B (2003) Face recognition in hyperspectral images. IEEE Trans Pattern Anal Mach Intell 25(12):1552–1560
14. Perona P, Malik J (1990) Scale space and edge detection using anisotropic diffusion. IEEE Trans Pattern Anal Mach Intell 12(7):629–639
15. Pietikinen M (2005) Image analysis with local binary patterns. In: Proceedings of Scandinavian conference on image analysis, pp 115–118

16. Roweis S, Saul L (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500):2323–2326
17. Socolinsky D, Wolff L, Neuheisel J, Eveland C (2001) Illumination invariant face recognition using thermal infrared imagery. In: Proceedings of IEEE CS conference on computer vision and pattern recognition (CVPR), vol 1, pp 527–534
18. Srivastana A, Liu X. Statistical hypothesis pruning for recognizing faces from infrared images. Image Vis Comput 21(7):651–661
19. Tan X, Triggs B (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. Trans Image Proc 19:1635–1650
20. Trujillo L, Olague G, Hammoud R, Hernandez B (2005) Automatic feature localization in thermal images for facial expression recognition. In: Proceedings of IEEE CS conference on computer vision pattern recognition (CVPR), vol 3, 14
21. Wolff L, Socolinsky D, Eveland C (2001) Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery. In: IEEE workshop on computer vision beyond the visible spectrum: methods and applications
22. Wu S, Song W, Jiang L, Xie S, Pan F, Yau W, Ranganath S (2005) Infrared face recognition by using blood perfusion data. In: International conference on audio and video-based biometric person authentication, pp 320–328
23. Xie Z, Wu S, Liu G, Fang Z (2009) Infrared face recognition based on blood perfusion and Fisher linear discrimination analysis. In: IST, pp 85–88
24. Zhili W (2002) Fingerprint recognition. University, Honk Kong Baptist

Chapter 7
Deep Feature Learning for Classification When Using Single Sensor Multi-wavelength Based Facial Recognition Systems in SWIR Band

Neeru Narang and Thirimachos Bourlai

Abstract In this chapter, we propose a convolutional neural network (CNN) based classification framework. Our proposed CNN framework is designed to automatically categorizes face data into individual wavelengths before the face recognition algorithms (pre-processing, feature extraction and matching) are used. Our main objective is to study the impact of classification of multi-wavelength images into individual wavelengths, when using a challenging single sensor multi-wavelength face database in short wavelength infrared (SWIR) band, for the purpose of improving heterogeneous face recognition in law enforcement and surveillance applications. Multi-wavelength database is composed of the face images captured at five different SWIR wavelengths ranging from 1150 nm to 1550 nm in increments of 100 nm. For classification based on CNN networks, there are no pre-trained multi-wavelength models available for our challenging SWIR datasets. To deal with this issue, we trained the models on our database and empirically optimized the model parameters (e.g. epoch and momentum) such that classification is performed more accurately. After classification, a set of face matching experiments is performed where a proposed face matching fusion approach is used indicating that, when fusion is supported by our classification framework, the rank-1 identification rate is significantly improved, namely when no classification is used. For example, face matching rank-1 identification accuracy, when using all data is 63% versus 80% when data is automatically classified into a face dataset where face images were captured at 1550 nm wavelength.

N. Narang · T. Bourlai (✉)
Lane Department of Computer Science and Electrical Engineering,
West Virginia University, Morgantown, WV 26506-6109, USA
e-mail: Thirimachos.Bourlai@mail.wvu.edu

N. Narang
e-mail: nneeru@mix.wvu.edu

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_7

Introduction

With the rise of global threats, security at the borders has become an area of great interest to prevent unlawful activities. One of the main challenges is to track an individual of interest and potentially identify her/him using their face images, utilizing as much information as possible from the collected database. Face recognition (FR) systems produce key trace evidence for the successful identification of potential threats. There are numerous challenges that operators have to mitigate, including the ability to detect and track humans when the face images are captured using different camera sensors, during day or night, behind glass, at different distances, indoors or outdoors etc.

One of the solutions to all possible challenges faced by operators is to design and develop an imaging system in the SWIR band. SWIR is a subset of the full infrared spectral range (in our experiments, the camera was sensitive from 0.9 μm to 1.7 μm) [1]. In harsh environmental conditions characterized by unfavorable lighting and pronounced shadows, FR based on SWIR images may be advantageous [2–5]. SWIR imagery is more tolerant to low levels of obscurants like fog and smoke [6–9]. The main benefit is that SWIR sensors can take advantage of sunlight, moonlight, or starlight [3, 10, 11]. With the use of SWIR sensors, we are able to capture images through tinted glass [12, 13].

Conventional imaging systems use a specific sensor (e.g. an SWIR camera) that can be operated without an external hardware, and utilize their complete spectral range to capture images. The information is collected over the wide spectrum and the integration process is responsible for getting less qualitative information than multi-spectral systems [7, 14, 15]. *Multi-Imaging Systems* (MIS) are either *Multi-Sensor* (MS), *Single-Sensor Multi-wavelength* (SSMW), or a combination of the two, i.e., *Multi-Sensor Multi-wavelength* (MSMW). MIS are composed of multiple sensors that operate in different bands. For example, we can use band-specific cameras to acquire images in the visible, NIR and SWIR bands. On the other hand, SSMW imaging systems utilize a single imaging sensor in combination with external hardware [16]. Such systems, before applying the aforementioned processing steps, are capable of acquiring images at specific wavelengths within the same band, e.g. a camera system like that can have a set of wavelength-selective band pass filters placed in front of the camera.

Motivation

Our designed and developed SSMW system [16] supports the acquisition and usage of unique facial information per individual that can enhance the performance of FR system. In our previous work [7, 16], an empirical optimization of the experimental set up was performed to acquire good quality face images. In this work, we are focusing on developing, the necessary algorithms to further pre-process the acquired

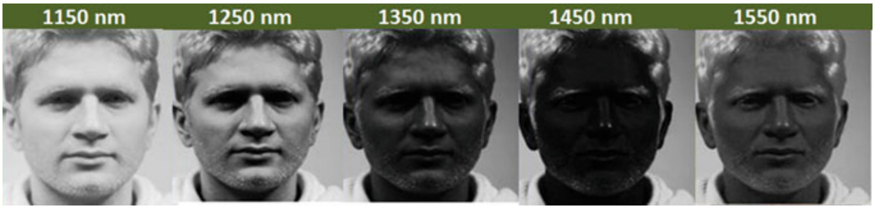


Fig. 7.1 Face images captured using the designed SSMW system operating in the SWIR band

face images. This design step is very important because it can further contribute to the performance of our proposed SSMW FR system.

Proposed CNN Network for Classification of Multi-wavelength Face Images into Individual Wavelengths

The face database is collected using our SSMW acquisition system. The system captured face images at 5 different wavelengths (1150, 1250, 1350, 1450, 1550 nm) in rapid succession, using a 5-filter rotating filter wheel as shown in Fig. 7.1. The major challenge when using a face database collected using our SSMW system operating in the SWIR band is, variation in image contrast and average brightness across the spectral bands as shown in Fig. 7.1. The spectral response (amount of absorbed and reflected light) is unique for each image at 1150, 1250, 1350, 1450 and 1550 nm. The present moisture content in the environment has an impact on the appearance of the face. Water vapors are responsible for more light energy absorption at 1450 nm causing face images to appear dark [7, 16]. However, the reflectance of light is greater for images at 1150 nm in comparison to the other four selected wavelengths. This difference in facial appearance can degrade the performance of the FR system including preprocessing, face detection, eye detection and face matching. To facilitate recognition performance, knowing the specific image category or image wavelength (based on the filter used) is important in order to set the proper parameters for image quality prediction, as well as face and eye detection.

Predicting this image category is a task that humans can perform easily. However, such a process is time consuming, especially, when using our SWIR database. This results in a large pool of images that need to be categorized to the right wavelength. Therefore, an automatic process of classifying images into individual wavelengths is needed. The automatic classification can be performed, using either supervised or un-supervised methods [17]. In this work, we propose a network, where the classification is performed using deep convolutional neural network. The main goal is to automatically classify the datasets into specific categories: multi-wavelength face images into individual wavelength for SSMW database, before FR algorithms are used. The big challenge we originally had to deal with was how to train the CNN model and, then, selection of hyper-parameters. The way we mitigated this challenge will be discussed in detail in the methodology section.

Proposed Heterogeneous Face Matching System

In a heterogeneous face recognition system, where the probe and gallery images are captured from different camera sensors, for example the probe images from SWIR or thermal cameras and the gallery images from visible camera. Traditional face recognition approaches, such as those based on Local Binary Patterns (LBP) or Linear Discriminant Analysis (LDA), often provide unsatisfactory results [18].

In this work, we propose a score-level fusion approach to perform the face matching experiments. It fuses scores from three descriptors LBP, Gabor and Histogram of Gradients-HOG (namely as LGHF operator: L for LBP, G for Gabor, H for HOG and F for fusion) based on the kernel subspace. The scores are computed when matching VIS to SWIR images, independent of the scenario they come from. Finally, a set of experiments are performed, based on individual feature descriptors including, LBP, Gabor Wavelets and HOG, and compared with the proposed face matching fusion approach.

The impact of classification of the multi-wavelength face database in terms of individual wavelengths is explored for the purpose of improving performance of cross-spectral multi-wavelength FR systems. Based on a set of experiments, we show that rank-1 identification accuracy results obtained from classification of the data and the proposed face matching approach, are higher across all the scenarios.

Background

For the identification of the wavelength of multi-wavelength images, in [19], the authors proposed the bag of features method to classify the images. This method generates a codebook or dictionary. This method was implemented using visible images to classify the objects based on a selected set of features. Namin et al. [20], proposed a method to classify visible and NIR images captured with a multispectral camera. The subjects' face images look similar when captured using a conventional camera but are easier to discriminate when captured using a multispectral camera. The classification was performed using local features and SVM classifier.

Recently, deep convolutional neural networks have achieved great success in the area of computer vision, machine vision, image processing and biometrics for the classification of scenes, object recognition, detection, face authentication and quality assessment. Some examples are the work of Gupta et al. [21] that proposed a probabilistic neural network (PNN) based approach for the classification of indoor versus outdoor visible band images. The segmentation is performed using C-means clustering and the features, i.e. color, shape and texture, are extracted from each image segment. In [22], the authors proposed a deep CNN for image classification. The authors selected the ImageNet database of over 15 million images, based on the classification results they reported that their CNN is capable of classifying highly challenging database. Levi et al. [23] proposed a deep learning method for gender and age classification for the database collected in un-constrained conditions in the visible band. The authors selected the Adience benchmark database collected from smart-phone devices. Based on the experimental results, they reported that their deep neural network can be used for the soft biometric classification.

This work is an effort to solve a more complicated problem as we are dealing with a multi-wavelength face database captured at variable pose angles in the SWIR band. We propose a deep learning based, sensor-adaptable algorithmic approach for the classification of data in terms of wavelength. What follows, is a discussion of our proposed methodological approach that manages to efficiently deal with a variety of problems related to the automated pre-processing and matching of multi-wavelength SWIR-based face images.

Methodology

In this section, we outline the databases selected to perform the experiments, a deep learning based method for the classification of multi-wavelength database in terms of individual wavelengths, and a set of experiments performed in order to find the significant impact of usage of classification of the data in face recognition.

West Virginia University SSMW (WVU-SSMW)

The constructed wheel, placed in front of the SWIR camera, has five filters, i.e. 1150, 1250, 1350, 1450 and 1550 nm (see Fig. 7.1). During data collection, we focused the camera at 1350 nm and collected the face images of 30 subjects [7]. The data collection was performed in 2 sessions on different days. Each session took 25 min, while both sessions required 50 min in total. In each session we collected 275 frames per subject (i.e. 5 set of selected wavelengths × 55 images per wavelength) [16]. The face images collected are left profile, right profile, and full frontal.

Classification Based on the Proposed CNN Framework

The classification of multi-wavelength SWIR images to individual wavelengths (from 1150 nm and up to 1550 nm in increments of 100 nm) is a challenging task. In our previous work [7], an automated quality-based score level fusion scheme was proposed for the classification of input multi-wavelength (MW) images. We selected features based on image quality factors such as sharpness, blurriness, structural content and contrast. The features were measured using reference and no-reference based quality assessment methods [24, 25]. In reference based image quality assessment method (average pixel distance, peak signal to noise ratio, mean square error etc.), the query image is compared against a reference image (collected under controlled conditions), and, the quality scores are generated [7]. In no-reference based method, no target images are used to establish the quality of the query image [26, 27]. The classification was performed using Bayesian and *kNN* models. We achieved greater than 96% accuracy. The key problem was the selection of features for the fusion, as the computed image quality scores were close to each other [7, 16]. The experiments were performed for within-subjects classification. To address these issues, we proposed a CNN based classification, where the selection of features is performed automatically. The experiments are performed for between-subjects classification. We compared the results from our proposed CNN network for within-subjects and between-subjects classification. To perform the classification, two databases collected in our lab namely WVU-SSMW and WVU Multi-Scenario (MS) are selected. WVU-MS database consists of 140 subjects and 13 samples for subject. The database consists of face images at five different wavelengths of 1150, 1250, 1350, 1450 and 1550 nm. WVU-SSMW database consists of face images from 30 subjects at 1150, 1250, 1350, 1450 and 1550 nm.

- **Model Architecture**: First images are rescaled to 32×32 and then input to the CNN network. In the first convolution layer, the input is filtered with 20 kernels of size 5×5, followed by a max pooling layer, which is taking the maximal value of 2×2. Max-pooling is a non-linear down-sampling operator. It divides the input image into a set of non-overlapping sub-image regions and each sub region outputs the maximum normalized gray scale value [28].

The previous layer is processed by a second convolution layer, with 50 kernels of size 5×5, followed by a max pooling layer. The second layer is further processed by a third convolution layer with 500 filters of size 4×4. This layer is followed by an activation function, i.e. a rectified linear unit (ReLU) [29]. Finally, a third layer is processed by the last convolution layer, with 26 filters of size 2×2 as shown in Fig. 7.2. The output of this layer is fed to a soft-max layer that assigns a label to each class, i.e. 1150 or 1250 or 1350 or 1450 or 1550 nm for WVU-SSMW database.

- **Training and Testing Data**: In the experiments performed, the subjects in the training and test sets are different, and the images are taken at different locations and days as shown in Fig. 7.3. To perform the classification, two sets of experiments are selected namely within-subjects and between-subjects classification (All Subjects). For within-subjects classification two scenarios are selected. *Scenario 1* is based on independent training and testing set. WVU-MS database (1150, 1250, 1350, 1450 and 1550 nm ground truth images collected indoors in SWIR band) is selected for training and WVU-SSMW database for the testing. For *Scenario 2*, the training data set is extended using WVU-SSMW database. WVU-MS and WVU-SSMW (20% of data) are selected to train the CNN network and rest of the WVU-SSMW database for the testing.

In between-subjects classification, for *Scenario 1* WVU-MS face images (with frontal view) are used for the training and WVU-SSMW database (with frontal and non-frontal view) for testing. For *Scenario 2*, we selected the WVU-MS database [30, 31] and extended the training data set using WVU-SSMW database (only 20% subjects). The rest of the data from WVU-SSMW database is selected for testing

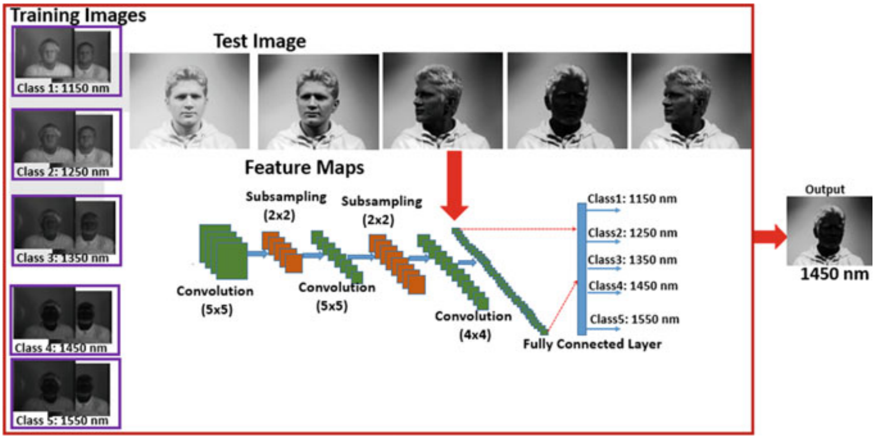


Fig. 7.2 An overview of our proposed CNN architecture

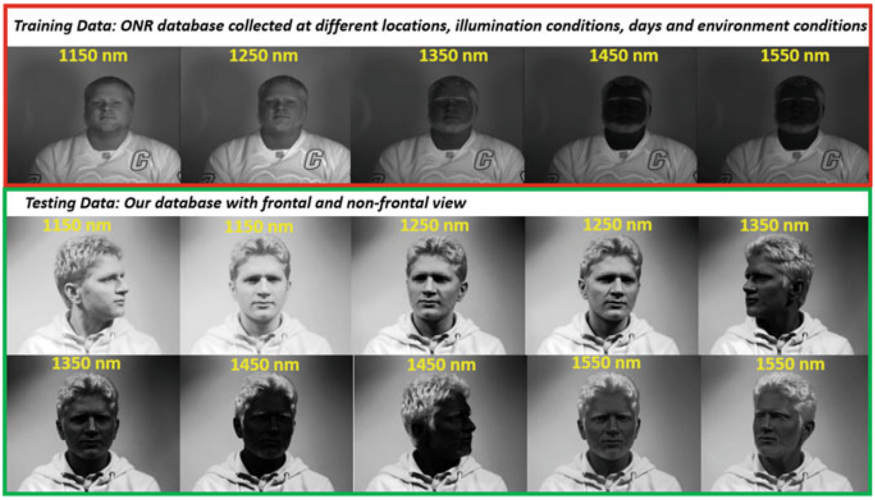

Training Data: ONR database collected at different locations, illumination conditions, days and environment conditions
1150 nm
1250 nm
1350 nm
1450 nm
1550 nm
Testing Data: Our database with frontal and non-frontal view
1150 nm
1150 nm
1250 nm
1250 nm
1350 nm
1350 nm
1450 nm
1450 nm
1550 nm
1550 nm

Fig. 7.3 Training and testing data for CNN architecture for classification

(no overlap of subjects). For *Scenario 3*, we selected the WVU-MS database [31] and extended the training data set using WVU-SSMW database (80% subjects). The rest of the data from WVU-SSMW database is selected for testing (no overlap of subjects).

Face Matching

In cross-scenario FR studies, we evaluate the system performance using academic and our proposed (LGHF operator based on fusion scheme) face matchers. We conducted cross-spectral (VIS vs. SWIR) face matching experiments. The main stages of face recognition system include, face normalization (neutral faces) and, then identification followed by feature extraction and matching.

- **Frontal and Non-Frontal View Face Classification**: WVU-MS database consists of face images with frontal view. While, for WVU-SSMW database, the face images consist of both the full frontal (FF) and non-frontal (NFF) view. The classification of the face images into frontal or non-frontal is performed based on our developed weighted quality-based score level fusion scheme in [7]. The quality scores used for the classification of FF versus NFF face images are computed based on reference and no-reference based quality assessment methods (such as the luminance, contrast, sharpness, blurriness).

- **Normalization of Data**: Image registration compensates for the slight perturbation in the frontal pose. It is composed of two main steps, eye detection and affine transformation [7]. The eye centers are, first, located by manual annotation and are used to geometrically normalize the images. Based on the located eye coordinates, the canonical faces are automatically constructed by applying an affine transformation. Faces are first aligned by placing the coordinates of the eyes in the same row, such that the slope between the right and left eye is zero degrees. Finally, all faces are canonicalized to the same dimension of 128 × 128 pixels.

- **Feature Extraction**: To extract the features, we selected the feature descriptors including i.e. LBP, Gabor and HOG [32, 33]. LBP descriptors are used to get the appearance and texture information and is invariant to change in illumination conditions [34]. It is highly discriminative, efficient method and perform well for the FR systems. Gabor wavelet is used to extract the shape and appearance information. Gabor wavelets transform provides desirable features and captures properties like spatial locality, spatial frequency and orientation selectivity. The features extracted from gabor filtered images are more robust to illumination and facial expressions. In this work, we used 40 gabor wavelets with the scale value ν lies in range from $\{0, 1, 2, 3, 4\}$ and eight orientations μ in range of $\{0, 1, -, 7\}$. *HOG* has proved as one of the successful local shape descriptor in computer vision. The main idea is to extract the local orientation information rather than the magnitude of image patches.

- **Matching based on proposed LGHF operator** In [18], we proposed a multi-feature scenario dependent fusion scheme to perform the face matching experiments in the near-infrared (NIR) band. Based on the results, we concluded that fusion based face matching scheme improves the performance in comparison to the face identification based on the individual feature descriptors. In this work, we extended our work to more challenging multi-wavelength where the face images are collected in the SWIR band with variation in pose angles.

The face images in our database vary in sensor type (VIS and SWIR) as well as in the pose. We selected kernel methods namely kernel linear discriminate analysis (KLDA) method to extract the discriminant information from a high dimensional feature space.

$$\Omega_{gallery} = P^T \phi(z_{gallery}) = (U\Lambda^{-1/2}V)^T k_{gallery} \tag{1}$$

$$\Omega_{probe} = P^T \phi(z_{probe}) = (U\Lambda^{-1/2}V)^T k_{probe} \tag{2}$$

$$d_{cos}(\Omega_{probe}, \Omega_{gallery}) = -\frac{\Omega^T_{probe}\Omega_{gallery}}{\|\Omega_{probe}\|\|\Omega_{gallery}\|} \tag{3}$$

For a given input face image (gallery/probe), its Gabor, HOG and LBP features are extracted and separately projected to the optimal discriminant feature space as illustrated in Eq. 1. In the projected space Eq. 1, Λ is the diagonal matrix of non-zero eigen values, U is the associated matrix of normalized eigenvectors and $k_{gallery}$ ϵ R^M is a kernel vector. The variable V, for kernel discriminative subspace, is computed by solving the LDA eigen decomposition. The training data is used to find the non-linear directions and Fischer's linear discriminant by mapping into non-linear feature space. In the testing set, for each subject in the probe set (SWIR band), we have a corresponding gallery image (VIS band). For each face image in the testing set (gallery and probe), first LBP, Gabor and HOG features are extracted and independently projected into the KLDA subspace. For the gallery image (VIS band), features are projected on to $\Omega_{gallery}$ feature space and for the probe face image features are projected to an optimal feature space Ω_{probe} (separately for each feature). The projected feature vectors are classified using nearest neighbor rule and cosine distance between the gallery and probe image as illustrated in Eq. 3.

The score normalization is performed using the *z-score* (score values lies between 0 and 1). The normalized scores are represented as z_{LBP}, z_{HOG} and z_{GABOR}. We propose an operator LGHF, where the fusion of the scores from LBP, gabor and HOG descriptors is based on decision level scenarios. To fuse the scores, 8 different scenarios are selected with different combinations namely, sum fusion scheme, we select three combinations: z_{GABOR}+z_{LBP}, z_{GABOR}+z_{HOG} and z_{HOG}+z_{LBP}. In the weighted fusion scheme, we select combinations with two features and, then, weights are assigned based on the performance of the individual feature descriptor. The rest of the two scenarios are based on all three feature descriptors, using the sum (z_{LBP}+z_{HOG}+z_{GABOR}) and weighted fusion scheme where weights are assigned to each descriptor based on the performance scores (distance scores).

Experimental Results

In this section, we aim to illustrate how the proposed classification system based on CNN network performs for multi-wavelength database. Our cross-spectral face matching results are discussed, when we utilize a set of feature descriptors (LBP, Gabor and HOG) and their fusion. Finally, we investigate whether our proposed approach improves performance when using classification of the data in terms of individual wavelengths for WVU-SSMW database. All these conditions are discussed in detail below.

WVU-SSMW Database Classification from Proposed CNN Architecture

The classification is performed for two sets including, within-subjects and between-subjects as presented in Sect. 1.3.2 (Training and Testing data). For between-subjects classification, we selected the momentum value of 0.92, batch size of 100 to train the CNN network. We conducted an empirical optimization on epoch parameter that resulted in better classification accuracy. We performed a series of experiments with the selected values of 4, 8, ..., 52. We repeated this process five different times for each epoch value using random selection of training and testing data (without overlap of subjects). In scenario 2 (between-subjects classification), based on mean and variance plots, the highest classification results are achieved for an epoch value of 44 as shown in Fig. 7.4 and the classification accuracy is more than 80%.

In Table 7.1, the classification results of highest accuracy are represented. For between-subjects, the classification accuracy is almost 35% for Scenario 1 (WVU-MS for training). For Scenario 2 (with extended training using WVU-SSMW database), the classification accuracy is more than 80% and more than 95% for Scenario 3. In order to examine the effectiveness of classification framework, each experiment is repeated 4 times for each cross-scenario (i.e. each time a different training set was

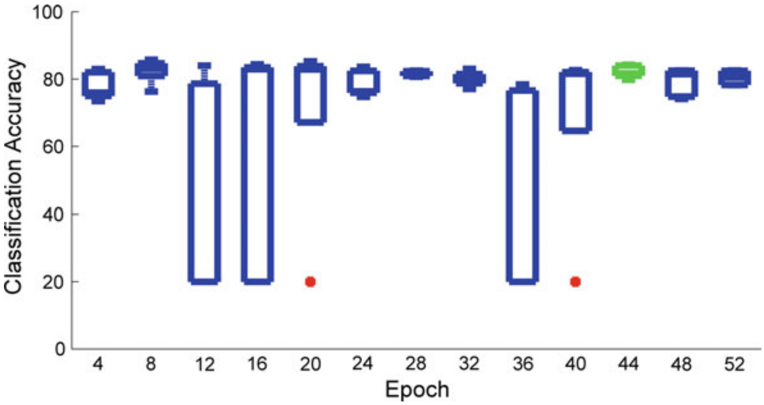


Fig. 7.4 Classification accuracy results with a selected set of epoch for CNN. Each boxplot is based on results from 5 randomly selected training and testing sets

Table 7.1 Classification results based on proposed CNN architecture

Classification accuracy	
Within-subjects	
Scenario 1	0.40
Scenario 2	0.94
Between-subjects	
Scenario 1	0.36
Scenario 2	0.84
Scenario 3	0.96

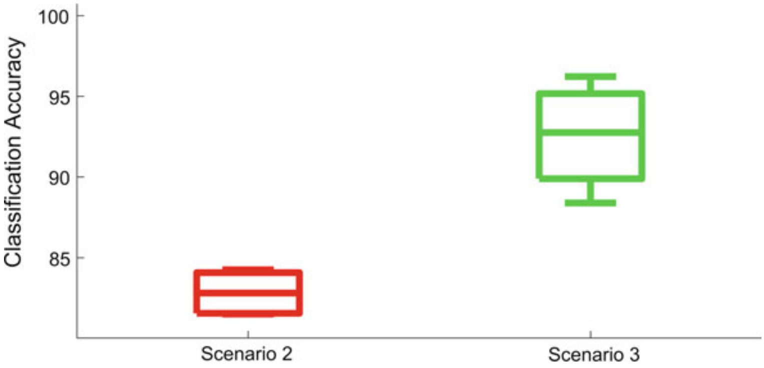


Fig. 7.5 Classification accuracy results for set 2: scenario 2 and scenario 3 with selected set of training sets for CNN. Each box-plot is based on results from 4 randomly selected training and testing sets

randomly selected) using proposed CNN network. Based on mean and variance plots as shown in Fig. 7.5, we concluded that the classification accuracy on average from 4 sets reaches greater than 80% for Scenario 2. For Scenario 3, the classification accuracy on average reaches more than 90%.

For within-subjects, the classification accuracy is almost 40% for Scenario 1 using the WVU-MS for training. For Scenario 2 with extended training using WVU-SSMW database, the classification accuracy is more than 90% as presented in Table 7.1. The system achieves promising classification results, for the extended training database. The decrease in the performance of the system for scenario 1 using only WVU-MS database for the training, is due to the variation in light conditions and sensors, under which the training images were collected. Moreover, the database consists face images with only frontal view. Whereas, for the WVU-SSMW database (testing data) face images collected are left profile, right profile, and full frontal. To address this problem, we extended the training data for example in scenario 2 and 3.

Cross-Spectral Face Matching: With and Without Classification of Wavelengths

Two sets of face recognition experiments are performed using the proposed fusion scheme based LGHF operator. First, experiments are performed with the original FR system, namely when no classification of the database is used. Without classification, all the data is used to perform the face matching experiments (VIS vs. All Data or Wavelengths). Second, experiments are performed to determine whether the classification of the data in terms of individual wavelengths (1150, 1250, 1350, 1450 and 1550 nm) can enhance the recognition performance. The classified database with labels; 1150 nm, 1250 nm, 1350 nm, 1450 nm, 1450 nm, 1550 nm from our developed deep learning system are used. For each dataset, we randomly divided the 50% data as the training set and rest of the data is used as the testing set, with no subject overlap.

The identification performance of the system is evaluated through the cumulative match characteristic (CMC) curve. The CMC curve measures the 1: m identification system performance, and judges the ranking capability of the identification system. For cross-scenarios, VIS versus SWIR face matching experiments are performed for 5 sets: VIS versus 1150 nm, VIS versus 1250 nm, VIS versus 1350 nm, VIS versus 1450 nm and VIS versus 1550 nm as shown in Fig. 7.6. This process is repeated five times, using random selection of the training and test sets each time, and, the rank identification accuracy results are represented in Table 7.2. The rank identification accuracy results for the first 5 ranks, for the best set are represented in Fig. 7.7. Each boxplot in Fig. 7.8 is based on rank-1 identification accuracy results from 5 randomly selected training and testing sets. Based on the mean and variance of boxplots, the matching accuracy from VIS versus 1550 nm, is overall higher compared to rest of the sets.

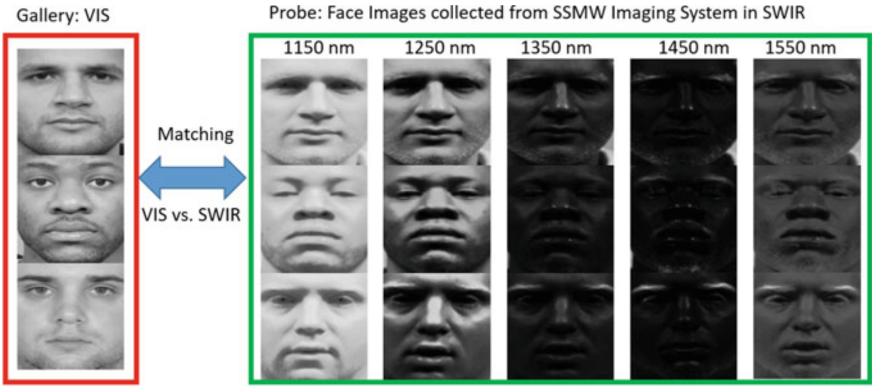


Fig. 7.6 Cross-spectral face matching: gallery images in the VIS band (Left) matched against the probe images in the SWIR band captured from our designed SSMW system at five wavelengths of 1150, 1250, 1350, 1450 and 1550 nm (Right)

Table 7.2 In this table we compared the Cross-spectral matching scenarios for VIS 1.5 m (Gallery) against SWIR (1150, 1250, 1350, 1450 and 1550 nm): experimental results when running our proposed FR algorithms using 50% of the data for training and the rest of the data for testing with no subject overlap. The experiments were run 5 times and the rank identification accuracy results presented here are the means

Gallery VIS versus SWIR

Rank	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
All Data (nm)	0.63	0.80	0.88	0.91	0.94
1150	0.74	0.83	0.87	0.90	0.91
1250	0.62	0.77	0.85	0.88	0.91
1350	0.74	0.91	0.97	1.00	1.00
1450	0.52	0.72	0.81	0.83	0.91
1550	0.80	0.88	0.93	0.96	0.97

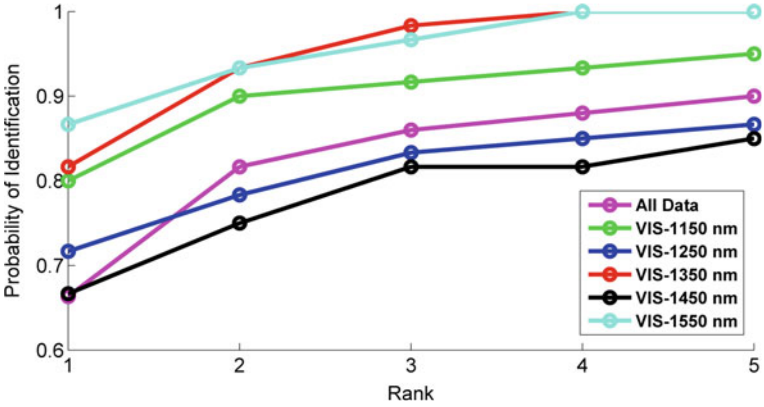


Fig. 7.7 Cross-spectral face matching scenarios for VIS (Gallery) against SWIR (Probe) for with and without classification of the data into individual wavelengths. The rank identification accuracy results are presented from best set

The CMC curves for the first 5 ranks of the best performed set out of randomly selected sets, are represented in Fig. 7.7. Based on the results, we determined that the rank-1 identification accuracy is improved from 66% (All vs. All) to 87% for the classified images into 1550 nm wavelength, to 80% for 1150 nm and to 82% for 1350 nm wavelengths as shown in Fig. 7.7. For class with label of 1250 nm wavelength, the rank-1 identification accuracy is improved from 66% to 70% but the rank-1 accuracy results are similar when selected the class with label of 1450 nm. However, based on rank-1 identification accuracy results from 5 randomly selected training and testing sets for 1450 nm class, the variance is large based on the box-plot representation as shown in Fig. 7.8 than rest of the classified classes. A possible reason for poor performance is the atmospheric absorption effect [7]. The present moisture content in the environment has an impact on the face appearance as described in Sect. 1.1. To

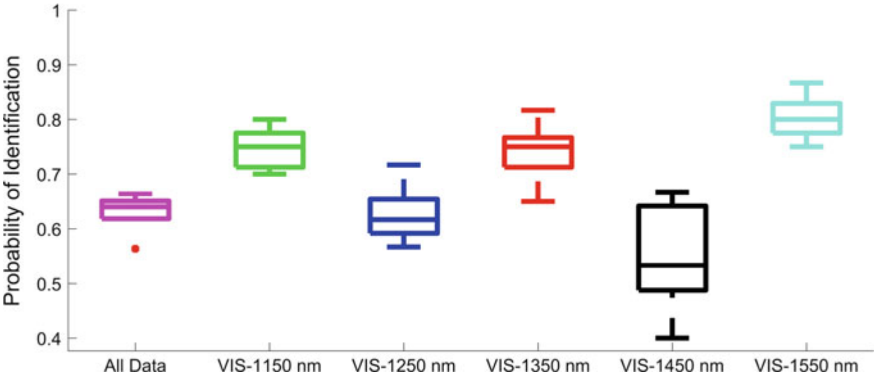


Fig. 7.8 Cross-spectral face matching scenarios for VIS (Gallery) against SWIR (Probe) for with and without classification of the data into individual wavelengths. Box-plots for cross-spectral face matching after running each experiment 5 times

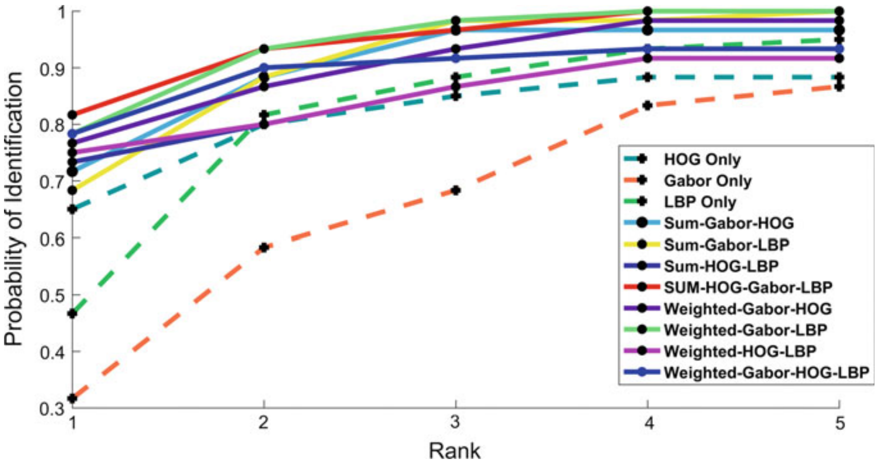


Fig. 7.9 Cross-spectral face matching scenarios for VIS (Gallery) against SWIR 1350 nm (Probe) for the data into individual wavelengths. CMC curves comparing the performance for individual versus Fusion based matchers

show the impact of our proposed fusion scheme (LGHF operator) to the identification accuracy of FR systems, we established a comparison between the individual descriptor and using the fusion schemes. The CMC curves for the first 5 ranks are represented in Figs. 7.9 and 7.10. Based on the results, it is concluded that the rank-1 identification accuracy is improved from 49% (LBP Only), 65% (HOG Only) to 82% (SUM-HOG-Gabor-LBP) for the probe images labeled with 1350 nm wavelength (see Fig. 7.9). For the images labeled with 1550 nm wavelength, the rank-1 identification accuracy is improved from 60% (LBP Only), 54% (Gabor Only) to 87% (Weighted-Gabor-HOG) (see Fig. 7.10).

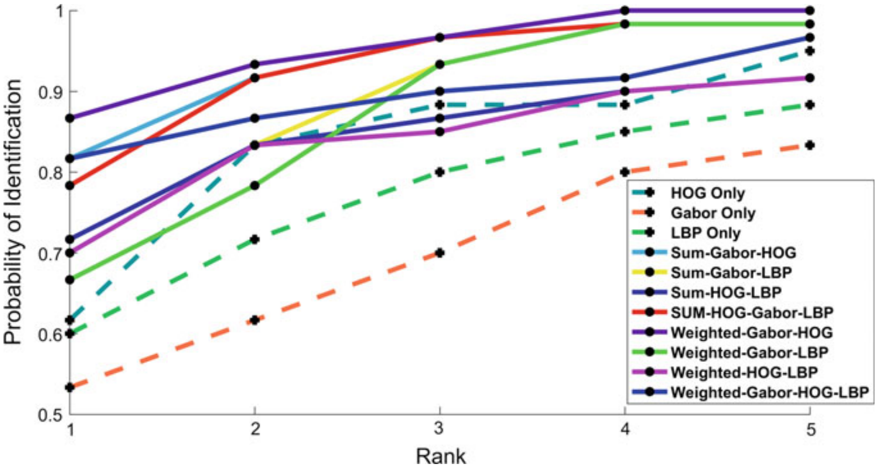


Fig. 7.10 Cross-spectral face matching scenarios for VIS (Gallery) against SWIR 1550 nm (Probe) for the data into individual wavelengths. CMC curves comparing the performance for individual versus Fusion based matchers

Conclusions and Future Work

In this work, we study the challenges of face recognition in the SWIR band i.e. when the face images are captured using multi-spectral imaging system against the visible (good quality) images. We proposed deep convolutional neural network based classification framework, designed to automatically categorize face data captured under various challenging conditions, before the FR algorithms (pre-processing, feature extraction and matching) are used. We trained the CNN model using our challenging SWIR face database and, for each classification level, a series of tests were performed to select the network parameters that result in high classification accuracy.

Our experiments showed that CNN provided us with significant classification accuracy, i.e. more than 90% for within-subjects classification and more than 80% for between-subjects classification (Scenario 2). Our face matching experiments provide important evidence that classification of the multi-wavelength database in terms of individual wavelengths provide significant improvement in the rank-1 identification accuracy, e.g., the performance is improved from 66% to 87% for the class with the 1550 nm label. Finally, we demonstrated that when using face images captured at 1550 nm, high identification rates are obtained. This conclusion is particularly important for unconstrained face recognition scenarios, where we may need to capture face images at 1550 nm (eye safe wavelength), at long distances and when operating at night time environments (preferable over other SWIR wavelengths).

Based on the experimental results we conclude that CNNs can be used to classify the data in terms of wavelength when using both constrained and unconstrained face datasets in the SWIR band. The classification of the datasets in terms of wavelength can provide significant improvement for face recognition in surveillance applications. In the future, we expect to further improve the classifications results. To do so, we plan to include more databases to train our deep learning model, as well as to test alternative CNN architectures.

Acknowledgements This work was sponsored in part through a grant from the Office of Naval Research (ONR), contract N00014-09-C-0495, Distribution A – Approved or Unlimited Distribution. We are also grateful to all faculty and students who assisted us with this work as well as the reviewers for the kind and constructive feedback. The views expressed in this chapter are those of the authors and do not reflect the official policy or position of the Department of Defense, or the U.S. Government.

References

1. Steiner H, Kolb A, Jung N (2016) Reliable face anti-spoofing using multispectral swir imaging. In: Biometrics (ICB), 2016 international conference on, IEEE, pp 1–8
2. Bertozzi M, Fedriga RI, Miron A, Reverchon JL (2013) Pedestrian detection in poor visibility conditions: would swir help?. In: International conference on image analysis and processing. pp 229–238
3. Kang J, Borkar A, Yeung A, Nong N, Smith M, Hayes M (2006) Short wavelength infrared face recognition for personalization. In: International conference on image processing ICIP, pp 2757–2760
4. Kong SG, Heo J, Abidi BR, Paik J, Abidi MA (2005) Recent advances in visual and infrared face recognitiona review. Comput Vis Image Underst 97(1):103–135
5. Lemoff BE, Martin RB, Sluch M, Kafka KM, McCormick W, Ice R (2013) Long-range night/day human identification using active-swir imaging. In: Proceedings SPIE infrared technology and applications XXXIX, vol 8704, pp 87,042J–87,042J–8
6. Martin RB, Sluch M, Kafka KM, Ice R, Lemoff BE (2013) Active-SWIR signatures for long-range night/day human detection and identification. In: SPIE defense, security, and sensing, international society for optics and photonics, pp 87,340J–87,340J
7. Narang N, Bourlai T (2015) Face recognition in the SWIR band when using single sensor multi-wavelength imaging systems. Image Vis Comput 33:26–43
8. Nicolo F, Schmid N (2012) Long range cross-spectral face recognition: matching SWIR against visible light images. Inf Forensics Secur IEEE Trans 7(6):1717–1726
9. Whitelam C, Bourlai T (2015b) On designing an unconstrained tri-band pupil detection system for human identification. Mach Vis Appl 26(7–8):1007–1025
10. Cao ZX, Schmid NA (2014) Recognition performance of cross-spectral periocular biometrics and partial face at short and long standoff distance. Open Trans Inf Process 1:20–32
11. DeCann B, Ross A, Dawson J (2013) Investigating gait recognition in the short-wave infrared (swir) spectrum: dataset and challenges. In: Proceedings SPIE biometric and surveillance technology for human and activity identification X, vol 8712, pp 87,120J–87,120J–16
12. Dawson J, Leffel S, Whitelam C, Bourlai T (2016) Collection of multispectral biometric data for cross-spectral identification applications. In: Face recognition across the imaging spectrum. Springer, pp 21–46
13. Ettenberg MH (2005) A little night vision-InGaAs shortwave infrared emerges as key complement to IR for military imaging. Adv Imag Fort Atkinson 20(3):29–33
14. Ngo HT, Ives RW, Matey JR, Dormo J, Rhoads M, Choi D (2009) Design and implementation of a multispectral iris capture system. In: Conference record of the forty-third asilomar conference on signals, Systems and Computers. pp 380–384
15. Steiner H, Sporrer S, Kolb A, Jung N (2015) Design of an active multispectral swir camera system for skin detection and face verification. J Sens 2016

16. Bourlai T, Narang N, Cukic B, Hornak L (2012) On designing a swir multi-wavelength facial-based acquisition system. In: Proceedings SPIE infrared technology and applications XXXVIII, vol 8353, Baltimore, USA, pp 83,530R–83,530R–14
17. Stallkamp J, Schlipsing M, Salmen J, Igel C (2012) Man vs computer: benchmarking machine learning algorithms for traffic sign recognition. Neural Networks J 32:323–332
18. Bourlai T, Mavridis N, Narang N (2016) On designing practical long range near infrared-based face recognition systems. Image Vis Comput 52:25–41
19. Nowak E, Jurie F, Triggs B (2006) Sampling strategies for bag-of-features image classification. In: European Conference on Computer Vision, Springer
20. Namin ST, Petersson L (2012) Classification of materials in natural scenes using multi-spectral images. In: IROS, IEEE, pp 1393–1398
21. Gupta L, Pathangay V, Patra A, Dyana A (2007) Das S (2006) Indoor vs Outdoor scene classification using probabilistic neural network. EURASIP J Adv Signal Proc 1:1–10
22. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Proc Syst 25:1097–1105
23. Levi G, Hassner T (2010) Age and gender classification using convolutional neural networks. In: Comput Vis Pattern Recognit CVPR Workshops
24. Caviedes J, Oberti F (2004) A new sharpness metric based on local kurtosis, edge and energy information. Signal Proc Image Commun 19(2):147–161
25. Wang Z, Bovik AC, Lu L (2002) Why is image quality assessment so difficult? In: Acoustics, speech, and signal processing (ICASSP), 2002 IEEE international conference on, IEEE, vol 4, pp IV–3313
26. Luxen M, Forstner W (2002) Characterizing image quality: blind estimation of the point spread function from a single image. Int Arch Photogrammetry Remote Sens Spat Inf Sci 34(3/A):205–210
27. Vu CT, Phan TD, Chandler DM (2012) bfS_3: a spectral and spatial measure of local perceived sharpness in natural images. IEEE Trans Image Proc 21(3):934–945
28. Vedaldi A, Lenc K (2015) MatConvNet-convolutional neural networks for MATLAB. In: Proceeding of the ACM international conference on multimedia
29. Parkhi OM, Vedaldi A, Zisserman A (2015) Deep face recognition. In: Proceedings of the British machine vision conference (BMVC)
30. Ice J, Narang N, Whitelam C, Kalka N, Hornak L, Dawson J, Bourlai T (2012) SWIR imaging for facial image capture through tinted materials. In: SPIE defense, security, and sensing, international society for optics and photonics, pp 83,530S–83,530S
31. Whitelam C, Bourlai T (2015a) Accurate eye localization in the short waved infrared spectrum through summation range filters. Comput Vis Image Underst 139:59–72
32. Cao D, Chen C, Piccirilli M, Adjeroh D, Bourlai T, Ross A (2011) Can facial metrology predict gender? In: IJCB, IEEE, pp 1–8
33. Tan X, Triggs B (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. Trans Img Proc 19(6):1635–1650
34. Cao Z, Schmid NA, Bourlai T (2016) Local operators and measures for heterogeneous face recognition. In: Face recognition across the imaging spectrum. Springer, pp 91–115

Part II
Surveillance for Security and Defense

Chapter 8
Managing Uncertainty from the Sky: Surveillance Through Three Generations of Commercial Earth Observation Satellites

Alexandros Kolovos

Abstract This chapter deals with the issue of surveillance and reconnaissance through space based earth observation commercial satellites, since they first appeared as a technology available to anyone forty five (45) years ago. Analysis divides this long period into "three generations of Earth Observation satellites". The first two generations, are distinguished from the third, using the criterion of derived products. The first two generations products were still imagery, while the third has added video too. The criterion of distinguishing the first and second generation, is the spatial resolution of their still imagery. As a result, the first provided general area surveillance, while the second provided reconnaissance. Implementing this criterion, these two first generations have lasted twenty six (26) years the first and fourteen (14) years the second accordingly. This chapter argues that the biggest breakthrough in decades actually occurred in 2014. A third generation has already breakout that year: it is the generation that has very high spatial resolutions in still imagery and videos too. This is also supplemented with high (to very high) temporal resolution (the time between image acquisitions). Minimizing the time between image acquisitions, is a key requirement for governmental users. It is this last technical characteristic which aims to cover the existing gap on the global quest (mainly from the users in the security and defence domain) for a permanent observation capability.

Introduction

This chapter deals with the issue of commercial earth observation (EO) satellites and their contribution to surveillance and reconnaissance from space. These satellites, which move with respect to the terrain beneath them, survey from their

A. Kolovos (✉)
Department of Automatic Control, Aerospace Technology,
Defense Systems and Operations, Hellenic Air Force Academy,
Dekelia Air Base, Dekelia, Greece
e-mail: alexandros.kolovos@hafa.haf.gr

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_8

low orbits above the surface of the earth the ground below. The satellites, which can serve, to various degrees, both commercial/civilian and military purposes firstly appeared as soon as the commercialization of space technology became available to anyone forty five (45) years ago.

This analysis divides this long period into "three generations of EO satellites". The first two generations are distinguished from the third, using the criterion of the derived products. The first two generations products were still imagery, while the third has added video too.

The criterion of distinguishing the first and second generation, is the spatial resolution of their still imagery. Resolution refers to the level of detail in the imagery collected by the satellite, the size of the smallest object one can detect with it [1]. The ability to distinguish the size and shape of an object of interest through high resolution image forming is a key element in intelligence derived from space.

Currently, high resolution in optical sensors is specified as capable of 3 m or less, while very high resolution (VHR) satellites are capable of resolutions well below of 1 m [2]. The analysis of this imagery can enable the precise description of important targets along with their accurate coordinates. Implementing this resolution criterion, these two first generations have lasted twenty six (26) years the first and fourteen (14) years the second accordingly.

We argue that the first commercial space EO generation provided general area surveillance, while the second generation permitted reconnaissance. In literature the term space observation is divided into two categories. The first is surveillance, which is the ability to detect changes in a country, which normally do not show any particular interest (e.g. the creation of a new port). Surveillance depends mainly in the spatial resolution of the systems. As shown in the table below, first generation systems gave general information only not full knowledge, they could just detect the existence of an object or not. The second category is reconnaissance. When surveillance detects changes in areas of interest (e.g. troop movements), at a time when there is tension or crisis in the two countries relations, then those who make decisions want to have more detailed information (Table 8.1).

This is done through "close looks", a more detailed and focused observation which requires frequent revisits, which ideally could lead to a permanent observation capability. Because of these characteristics, satellites are a part of the Intelligence, Surveillance and reconnaissance (ISR) capabilities which collect, process and exploit accurate and timely information that can provide the battlespace awareness.

This chapter argues that the biggest breakthrough in decades actually occurred in 2014. A third generation has already breakout that year: it is the generation that has very high spatial resolutions in still imagery and videos. This will be also supplemented for the first time with high (to very high) temporal resolution (the revisiting frequency, the time between image acquisitions over a specific location) which gives greater capacity, along with a greater spectral diversity [1]. This capability is a critical parameter since it is needed especially in periods of crisis, because crisis depends on revisit time.

Table 8.1 Spatial resolution requirements for surveillance and reconnaissance

Target	Detection[a]	General id[b]	Precise id[c]	Description[d]	Technical analysis[e]
Bridges	6	4.5	1.5	1	0.3
Radar	3	1	0.3	0.15	0.015
Supply dumps	1.5–3.0	0.6	0.3	0.03	0.03
Troop units (in bivouac or on roads)	6	2	1.2	0.3	0.15
Airfield facilities	6	4.5	3	0.3	0.15
Rockets and Artillery	1	0.6	0.15	0.05	0.045
Aircraft	4.5	1.5	1	0.15	0.045
Command and control HQ	3	1.5	1	0.15	0.09
Missiles (SSM/SAM)	3	1.5	0.6	0.3	0.045
Surface ships	7.5–15	4.5	0.6	0.3	0.045
Nuclear weapons components	2.5	1.5	0.3	0.03	0.015
Vehicles	1.5	0.6	0.3	0.06	0.045
Minefields (land)	3–9	6	1	0.03	0.09
Ports and harbors	30	15	6	3	0.3
Coasts and landing beaches	15–30	4.5	3	1.5	0.15
Railroad yards and shops	15–30	15	6	1.5	0.4
Roads	6–9	6	1.8	0.6	0.4
Urban areas	60	30	3	3	0.75
Terrain	–	90	4.5	1.5	0.75
Submarines (surfaced)	7.5–30	4.5–6	1.5	1	0.03

Sources Senate Committee on Commerce, Science, and Transportation, *NASA Authorization for Fiscal Year 1978,* 1642–43; and *Reconnaissance Handy Book for the Tactical Reconnaissance Specialist* (St. Louis, Mo.: McDonnell Douglas Corporation, 1982), 125

[a] Detection: Location of a class of units, objects, or activity of military interest

[b] General Identification: Determination of general target type

[c] Precise Identification: Discrimination with target type of known types

[d] Description: Size/dimension, configuration/layout, component construction, equipment count, etc

[e] Technical Analysis: Detailed analysis of specific equipment

Since satellites in low orbits move fast with respect to the earth beneath them, they can collect imagery for some minutes before they move along their track. The next time they pass after 90 min, they do not see the same area, due to earth's move and one has to wait some days before the same scene can be revisited. This creates coverage gaps, which have profound implications in the case of intelligence collection. Minimizing the time between image acquisitions, is a key requirement for governmental users.

One way to do this is to put a large number of satellites in orbit simultaneously, so that the whole earth is continuously monitored. It is this last technical characteristic which aims to cover the existing gap on the global quest (mainly from the users in the security and defence domain) for a permanent observation capability. Other ways, which may include i.e. geostationary satellites or even systems in near-space that are meant to fly for months or even years like stratospheric balloons, pseudo-satellites and high-altitude drones fell outside the scope of this chapter.

The commercial satellite imagery business consists mainly of optical imagery. These are complemented by RADAR systems such as Canada's MacDonald Dettwiler and Associates (MDA) Radarsat-2, Cosmo-SkyMed from Agenzia Spaziale Italiana, RapidEye's 5 satellites from Germany's DLR, and TerraSAR-X, a public-private-partnership between the DLR and EADS Astrium (now Airbus Defence and Space). This all weather and all day/night capability is required for a permanent observation capability.

This last generation is driven not only by the usual operational requirement to have a continuous basis (24 h per day) observation capability. It is also helped by the advances in technology, which permits enhanced capabilities in large constellations of smaller satellites, which permit unprecedented revisit time, with cheaper costs. This third generation will be a stepping stone closer to the quest of a "total information awareness". This article tries to raise the awareness of the public of what it is expected to come the next few years, if the life cycle of this third generation continues to be smaller, as the first two had.

Defining Three Generations of Commercial Earth Observation Satellites

a. The first generation (1972–1998): Poor Man's Surveillance

Looking back, at the time when the first generation of commercial earth observation satellites was launched in 1972, one can easily identify that space capabilities were shared among a small group of major space-faring actors. These were capable of production, maintenance, and control of satellites and launchers. After all it was still the Cold War and commercial space activity was not that significant.

The United States was first to launch satellites remotely sensing the Earth but did so within the defence sector. Then the decision to commercialize this technology came. American first U.S. multispectral satellite ERTS (later renamed to Landsat-1), created the breakout showing to the public some of the capabilities a small space community enjoyed until then. Designed to last only 3–5 years, Landsat presented for the first time to the public some of the privileges that only US and USSR had at that time. But, it's spatial resolution was poor. One could see mainly synoptic views of our planet. With a resolution of 80 m and a broad area coverage of 170 × 185 km, one could just recognize big airports and ports in its still imagery. This resolution has been restricted due to political reasons and not technology barriers. At that time US military reconnaissance satellites had spatial resolutions below half a meter using school bus-sized satellites. Also its temporal resolution was low: it had a revisit time of 18 days [27].

Landsat-1 and its successors held a full global monopoly for 14 years, until Europe's reply came with the launch of the French satellite SPOT-1 (1986). The spatial resolution improvement was huge. From the 30 m spatial resolution the Landsat-4 had, SPOT1 gave to the public a seriously improved capability with 10 m (but with a smaller 60 × 60 km area coverage), which lasted alone for almost ten whole years. The SPOT satellite could take an image of an area every 26 days. But with an oblique viewing capability, its revisit time could fall to 5 days (varies with latitude), while it offered stereo pairs for the first time too. SPOT-1 was the ideal tool at that time for surveillance, as its company advertised [31]. On the other hand it had a just acceptable georeferenced horizontal accuracy of 350 m. Also the delivery of its imagery took weeks or even months.

But even with these poor characteristics, there was some use for surveillance purposes. A classic example is the accidental detection of the Soviet Krasnoyarsk RADAR in 1983 by a Landsat satellite. This development, which violated the bilateral U.S.—U.S.S.R Antiballistic Missile (ABM) Treaty, had been in progress from 1979. But it had not been detected by the dedicated U.S. military satellites which were not tasked to observe that remote area in Siberia, away from soviet borders [21].

It was around the end of the Cold War that Russia decided to export its reconnaissance products, held until that time only for internal use. It was around 1988, when the first KVR 1000 camera's color scenes appeared in the west, with a spatial resolution of around 2 m. These images broke the high resolution monopoly the photoreconnaissance satellites of the U.S. and the USSR had for 25 years and showed everyone what imagery from space could reveal.

The importance of the above civilian satellite systems was also highlighted in the first space war, the Gulf war in 1991. There, the U.S. LANDSAT and French SPOT civilian systems were for the first time used by the U.S. military to supplement its military reconnaissance satellites to accomplish military goals [20].

From that time, the commercial competition started to catch some speed. In 1995, India launched IRS-1, with 5.8 m spatial resolution, a revisit of 5 days and an area coverage of 70 × 70 km scenes. Although it seemed like a clear winner in the commercial earth observation arena in clarity temps, IRS had some drawbacks, mainly in terms of the accuracy of its images. To keep costs and weights down, IRS was not designed to offer the best direct georeferencing performance (circa 800 m), an issue that is of special interest to security and defence users.

Optical satellites were the dominant systems in this first generation, while EO radar data market was at an infant stage. These first generation systems, provided a general surveillance capability for the public, with a few satellite systems which possessed imagers with low to medium resolutions. It lasted twenty six (26) years.

Although the American Landsat was the first to start the commercial space race, US response to the strong competence from European, Indian and Israeli space systems was hectic at least. The U.S. through a series of Presidential Directives (see [3, 12, 16, 28], finally decided to lift the political barriers it had imposed two decades ago to the permitted level of spatial resolution. Thus, the next generation was led by US optical satellites presenting new capabilities with spatial resolutions below 1 m.

b. The second generation (1999–2013): The boom in GEOINT Capability (reconnaissance)

The second generation of commercial Earth Observations satellites started back in September 24, 1999 with the long anticipated launch of Ikonos-1 (the name comes from the Greek word for image. One of its ground receiving stations originally had been in Greece).

That system historically holds the title of the first satellite which broke the limit of 1 m resolution, with its 0.82 m panchromatic spatial resolution [30]. Also its revisit time was better, it was approximately 3 days. But at the same time, this very high resolution resulted in a narrower field of view (with an 11 x 11 km scene coverage). Ikonos was not only the first "very high spatial resolution" satellite, but it still holds the record of longevity of this second generation. It stayed 16 years in low orbit, before decommissioning in 2015.

Afterwards, the commercial proliferation of space technology has radically increased. The competition race of the earth observation satellite got speed. Series of various sophisticated systems were build and operated, not only from major space-faring actors but from commercial companies too, offering new unheard so far services which promoted autonomy and secrecy.

First it was the Israeli Earth Remote Observation System A (EROS-A) as a part of the EROS family, by ImageSat International with a resolution of 1.8 m. It was launched in 2000 and one of its competitive advantages was that it could sell exclusive rights to directly task the satellite during its passage over a specific area without anyone else could do this by the same system [5].

But since the US mentality had already changed, rapidly came the reply. New, more advanced dual-use satellites, like Quick Bird 1 and 2 quickly emerged in 2000 and 2001. Their key characteristic was that they had very high spatial resolutions, they were smaller than the previous large satellites, and had relatively good revisit rates (1–3.5 days depending on latitude). They were also cheaper and had a longer life expectancy of 8–12 years [23].

Israeli EROS-A was followed by EROS-B with a higher resolution of 0.5 m panchromatic resolution at 500 km altitude, in 2006. Then after them US GeoEye (2007), and WorldView (2007) appeared quickly.

These advanced private satellites orbiting the Earth, were competitors at a time when the 2007–2008 global financial crisis hit. Obviously this had an impact on their businesses. Since the US Intelligence Community uses also private satellite imaging firms to conduct national security and defense missions and the mentality had changed, instead of letting these private-sector companies go bust, they were augmented.

This was a time when the US intelligence had specific gaps in their spy satellites so it was necessary to supplement some capabilities from the commercial sector [24]. Various partnerships were formed between the US. Government (specifically from DOD's National Geospatial Intelligence Agency-NGA) and the commercial satellite industry. After a series of negotiations and buys, all these systems, along with Ikonos fell under the ownership of DigitalGlobe, which currently is the world's leading provider of commercial satellite imaging.

In the meantime, France replied with a move to the use of smaller satellites (with a weight of about 1000 kg instead of 3 tons as for the SPOT-5). This resulted to a new French generation of optical imaging satellites, the Pleiades 1A and 1B, launched in 2011 and 2012 accordingly. Its spatial resolution is 0.70 m (after resampling 0.50 m) and its location accuracy, without ground control points is around 3 m (CE90). They have a daily revisit capability.

All these systems, along other from smaller actors, provide very high resolution still imagery along with high accuracy in georeferencing capability. These two enhanced characteristics permitted the move from the generic surveillance phase, which provided imagery intelligence (IMINT) to the Geospatial Intelligence phase. GEOINT integrates imagery, imagery intelligence and geospatial information (Fig. 8.1).

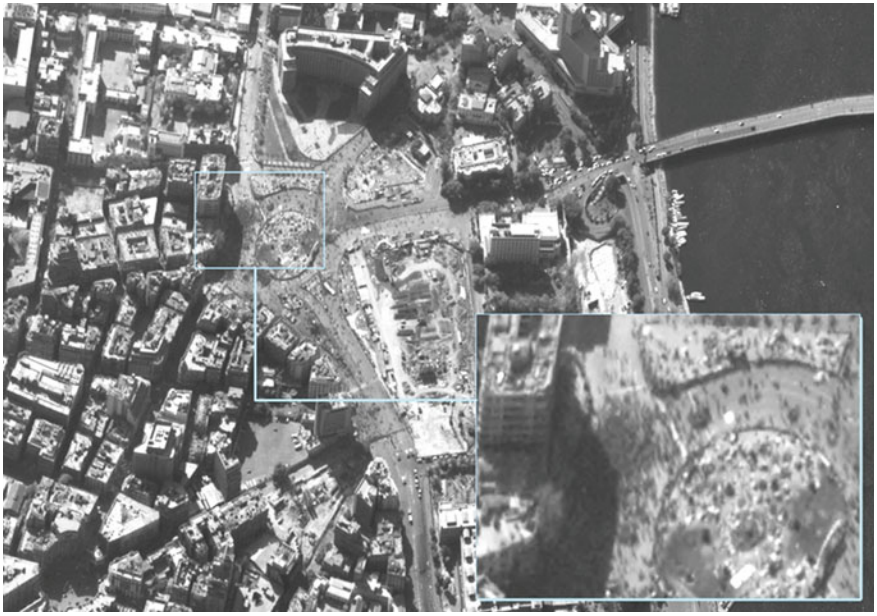


Fig. 8.1 Monitoring the "Arab Spring" uprisings via commercial 2nd generation satellites. People gathering on a square during a crisis. *Source* EU SatCen, https://www.satcen.europa.eu/services/humanitarian_aid. Accessed July 24, 2017

Also, before the end of this second generation the transition from still imagery to videos has been demonstrated. The SkySat-1 satellite, with a size of 60 × 60 95 cm and weight of approximately 100 kg, was launched by Skybox Imaging on November 21, 2013. It was the first commercial satellite which produced a video for up to 1.5 min. The stage for the next generation has been again set by the US, as it was ready to soon add satellites capable of 0.3 m resolution.

c. The third generation (2014–): The trend towards very High temporal resolution

The trend that drove the second generation was the drive to be small, efficient and inexpensive than relying on a handful of big satellites that are difficult to replace.

This third generation is full of breakthroughs and innovations. They focus on many constellations of even smaller and cheaper satellites, with higher than before, temporal resolution. It is characterized by an increased number of space actors, which now include a number of non-state actors.

The latter have developed and orbited or are developing new sophisticated systems. They not only have high (to very high) resolutions but have also the ability to look at any spot on the planet multiple times per day. Thus, they are capable of rapid acquisitions of still imagery and short videos of the earth's terrain. But, currently the space industry is a scattered and still fluid situation, with many brands, and no single company has been able to really put the whole package together.

2014, according to this analysis is the first year of the third generation of commercial earth observation satellites. Although new systems with enhanced capabilities started to be launched in late 2013 (Skybox-1, which provided High Definition (HD) videos, it is the following selected main events that made us consider 2014 as a pivotal year:

1. In late 2013 and in 2014 the US based startup company, SkyBox launched the SkySats 1 and 2, with a 0.90 m panchromatic spatial resolution and multi-spectral imagery. Also it provided the first ever commercial High-Definition full-frame rate panchromatic video to be captured from space, with approximately 1 m resolution [19]. In 2014 it was bought by Google for US$ 500 million and renamed as TerraBella. The latter in 2016 launched five (5) satellites in 2016 (one in June 2016, and four launched simultaneously in July 2016). Thus TerraBella is the first multitemporal high resolution constellation with its seven (7) SkySat satellites. These satellites weigh around 120 kg, so they are tiny compared to most commercial satellites. Each satellite will offer still and video products and will be capable of sub-meter resolution (but lower resolution than DigitalGlobe's satellites or French Pleiades). To keep costs and weights down, Skysats were not designed to offer the best direct georeferencing performance (around 100 m circular error). The company plans to have its fleet of 21 satellites in orbit by the end of 2017, so a revisit rate of roughly three times a day for most of the planet, could be succeeded. If this materializes, this would be the first time the Earth would be seen at so many different times per day and

night. This will have obvious implications for the security and defence end users toward the quest for a permanent earth observation capability.

2. Between 2013 and 2017, Planet, a U.S. based startup company, has placed 190 “CubeSats” into orbit which completes its initial fleet of spacecraft, on 14 different launches. This is currently the largest constellation in orbit. Planet builds large quantities of cheap satellites, called ‘Doves’. Their weight is about 5 kg and can take images at an average resolution of 3.7 m. Instead of picking the highest resolution it focuses on a high temporal cadence to see things every day, independently of tasking. Although limited by their high resolution, Planet insists that they will be able to image the entire planet in one day, every single day. In the future, Planet plans to be anywhere, anytime by imaging every location on earth, every day, at high resolutions, below 1 m, while it also experiments with near-infrared. The Planet team does not custom design every piece, but uses common parts from the consumer electronics market to build satellites. In 2015, Planet acquired another satellite company, Black bridge, which operates a five satellite constellation of 5 m resolution EO “RapidEye” satellites, build by the British Surrey Satellite Technology LTD. In early 2017, Planet set a world record by launching 88 of its Doves aboard an Indian Polar Satellite Launch Vehicle (PSLV) rocket, which consists so far the largest fleet of satellites launched in history [26]. On July 14, 2017 Planet launched another set of 48 satellites from a Russian Soyuz rocket (Fig. 8.2).
3. In January 2014, Urthecast, based in Canada, began by operating 5 and 1 m resolution imagers, on the International Space Station (ISS) [6]. UrtheCast was the second company to offer video from space. A camera called Iris, can acquire the first Ultra-High Definition (UHD) videos (60 s long), over any location that

Fig. 8.2 Planet’s bread loaf-sized satellites, called ‘Doves’. *Source* https://www.planet.com/. Accessed July 8, 2017

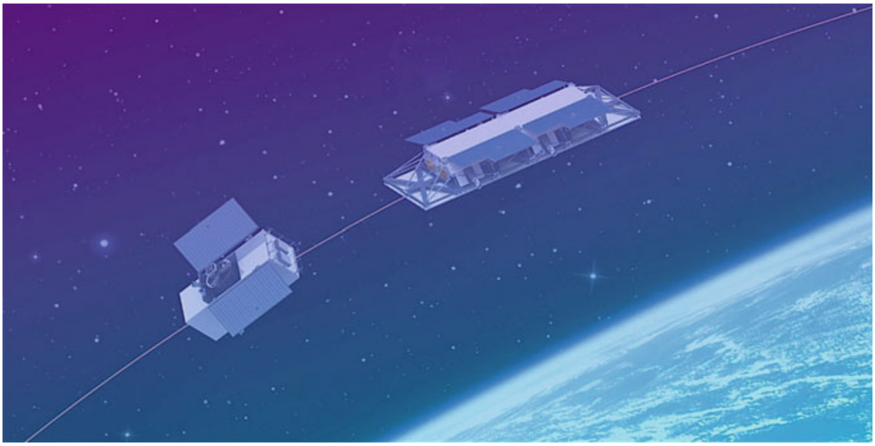


Fig. 8.3 Urthecast OptiSAR constellation's tandem pairings are designed to allow for near-simultaneous acquisition of SAR and optical data. *Source* https://www.urthecast.com/optisar/. Accessed July 23, 2017

the space station orbits [13]. Video has color, and covers a wide area. But still, these imagers do not possess the required geolocation accuracy for security users [6]. Urthecast is now in the process of creating its own OptiSAR multi-spectral optical and Synthetic Aperture Radar (SAR) constellation of 16 Earth Observation satellites. Constellation is consisting of eight optical and eight SAR (Synthetic Aperture Radar) satellites in LEO, which is expected to provide frequent still images and video up to 0.50 m resolution (Fig. 8.3).

4. In 2014–2016, the U.S.-based DigitalGlobe Inc, currently the global leader in commercial EO satellites, launched the two most capable satellites in terms of spatial resolution and precision of geo-location accuracy. World View-3 satellite, launched on August 13, 2014, has 29 spectral bands (spectral resolution specifies the number of spectral bands in which the sensor can collect reflected radiance) including a panchromatic channel with 0.31 m ground resolution. The satellite has an average revisit time of <1 day and much of its products has been prepurchased for US governmental uses. Then in November 2016, Worldview 4 followed. Its commercial available imagery is also limited to 0.31 m spatial resolution (Fig. 8.4).

   It is estimated that its real resolution is even higher, closer to 0.20 m, but restrictive licensing by the U.S. government makes it likely that only the US government will have access to imagery at the full design resolution. Currently DigitalGlobe has five (5) satellites in orbit. Combined all 5 satellites can revisit any place on the planet three or four times every day. With weights of 2–2.5 tons, they are fully operational and paid for and require only ongoing costs to maintain this constellation. One of DigitalGlobe's most valuable assets is its vast imagery archive, which is growing daily. These big data are estimated to cover 7 billion sq. km. totaling 100 petabytes of data [9]. Apart from this, DigitalGlobe

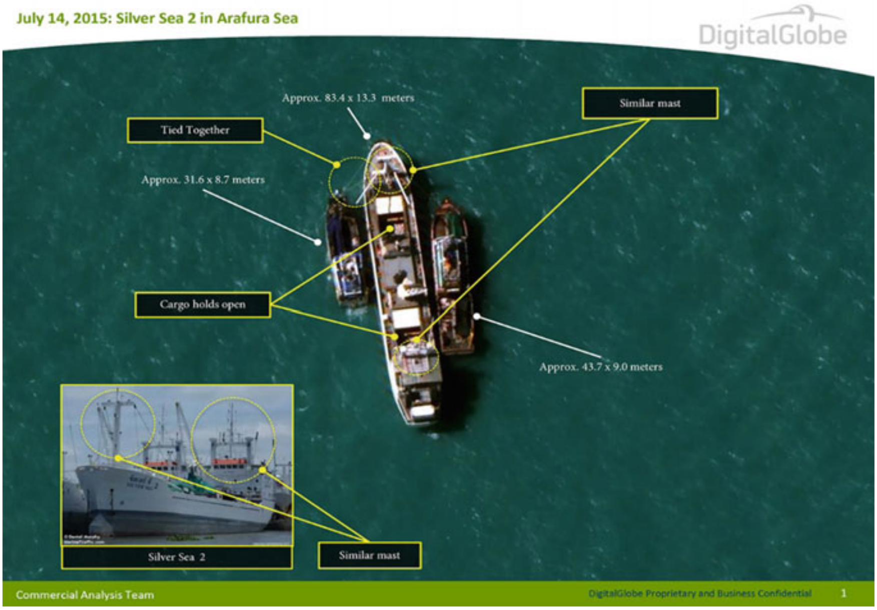

July 14, 2015: Silver Sea 2 in Arafura Sea
DigitalGlobe
Approx. 83.4 x 13.3 meters
Tied Together
Similar mast
Approx. 31.6 x 8.7 meters
Cargo holds open
Approx. 43.7 x 9.0 meters
Silver Sea 2
Similar mast
Commercial Analysis Team
DigitalGlobe Proprietary and Business Confidential
1

Fig. 8.4 Slavery from space: locating and identifying fishing vessels engaged in illegal activity by delivering indisputable evidence of human trafficking. More than 2,000 slaves were freed. *Source* DigitalGlobe2 [18]. Accessed July 23, 2017

will launch a new constellation in 2019 of six smaller satellites, together with two Saudi companies, each with a ground resolution sharper than 1 m. Also it will launch its own bigger satellites in 2021. In conventional terms, DigitalGlobe remains the standard in the VHR commercial space market. It responds to customer tasking with products which balance well between a very high spatial resolution and a high revisit time which permits daily global coverage.

5. In 2014–2015 the Korea Aerospace Research Institute (KARI) has launched two optical very high-resolution Korean Multi-purpose Satellites (KOMPSATs 3 and 3A). Their spatial resolution of 0.70 and 0.40 m accordingly, brings the latter as the second best very high resolution of the world today. KARI initially developed KOMPSAT, a small 500 kg Earth observation satellite with an orbital altitude of 685 km, jointly with U.S. TRW Inc. (which has built also Signal Intelligence spy satellites). KARI has plans for 10 more satellites in the pipeline for the coming decade (Fig. 8.5).
6. In 2014, Argentinean Company Satellogic S.A launched its first technology demonstration mission, with a long-term goal of having 300 satellites by early next decade. At that stage customers will supposedly be able to get an image in 5 min at 1 m resolution. Satellogic satellites will also carry hyper spectral imaging, with 30 m spatial resolution. These satellites are built with newer electronics technology and have the size of "a desktop computer hard drive"

Fig. 8.5 KOMPSATs family of satellites. *Source* https://www.blacksky.com. Accessed April 21, 2017

[17]. Satellogic has strong support from the Chinese Tencent (which is among the 10 world's largest companies and uses of the Long March launcher. Currently Satellogic has three high-resolution satellites and expects this number to grow to a constellation of over 60 satellites by 2019 (Fig. 8.6).

7. In 2016 the launch of the first Pathfinder-1 of the U.S. based BlackSky Global paves the way for its 60-satellite constellation by 2020. The first three commercially operational satellites, with 1 m high-resolution are scheduled for launch in 2017. They have approximately the size of a mini-refrigerator with weight of 50 kg. Their geolocation accuracy is very good and is smaller than 10 m (CE90). The complete constellation is scheduled to be on orbit by 2020. The satellites will replaced every 3 years. For $90, people can receive images (4.4 × 6.6 km/29 km^2) within 90 min of landscapes across the world. In the past delivery took days or weeks and the products had higher prices [11]. When Ikonos had the very high resolution monopoly, circa 2000, its 1 m resolution imagery costed $38 per sq km.

The case presented above, in my judgement support the argument that we have entered a new phase of proliferation in the commercial space earth observation domain. The main driver for the third generation, are the enhancements in micro technology, which permits the existence of constellations of smaller satellites with some great capabilities. For example in terms of weight, each Skybox satellite only weights around 120 k (265 lb), while each so it is tiny compared to most commercial (but also more capable) satellites like the DigitalGlobe satellites which weight 2 tons each.

This affects also the cost of each satellite. In the U.S. the trend of modernizing the nation's aging satellite-imagery architecture by enhancing use of U.S. commercial providers has begun almost 20 years ago. This move which relies also on private-sector companies as the main source of satellite imagery had a target to save the U.S. government billions of dollars. But these companies are subject to scrutiny by the U.S. government to ensure that sensitive information is not released to unauthorised parties (Fig. 8.7).

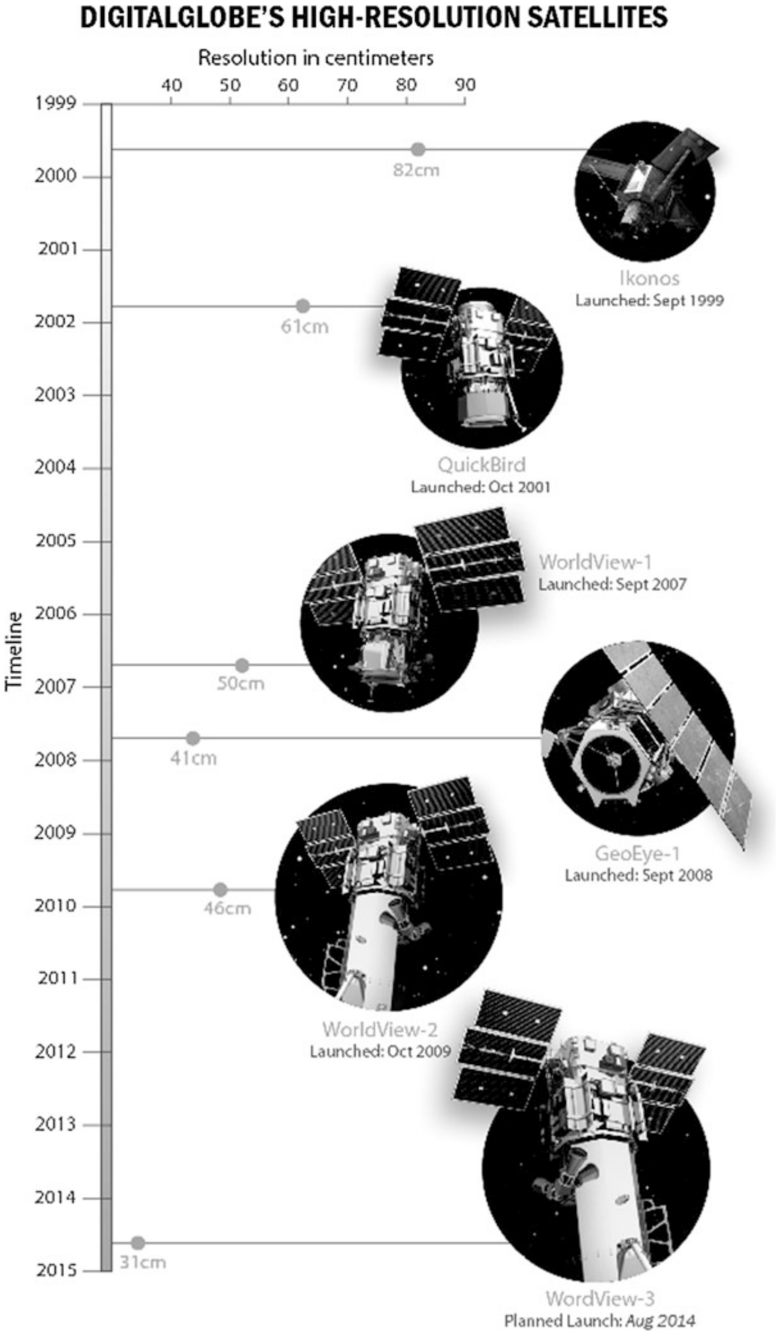


Fig. 8.6 DigitalGlobe's very high resolution satellites. Accessed 2 April, 2014

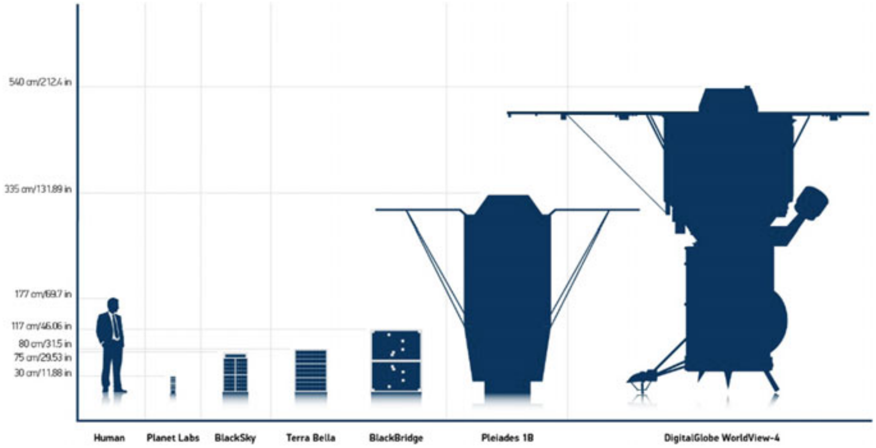


Fig. 8.7 Size comparison between second and third generation commercial EO satellites. *Source* DigitalGlobe.com. http://blog.digitalglobe.com/news/frequently-asked-questions-about-worldview-4/. Accessed 24 July, 2017

Military satellites are extremely costly to build, launch and operate. The KH-11 was first launched in December 1976 to serve close-look and area surveillance purposes on a single satellite. It also presented for the first time digital imagery technology, which revolutionized time availability of the imagery by returning them virtually instantaneously via a relay satellite and presented an entirely new method of interpretation [7]. The letters "KH" are said to stand for "Key Hole". Exact costs are classified, but estimates regarding the cost of a current US military optical reconnaissance satellite KH-11 range between $4–6 billion per satellite [25].

Each DigitalGlobe's satellite costs about $600 million [8], so each DigitalGlobe' satellite cost is around one tenth of the cost of a highly classified KH-11. In these costs one should add the cost to launch the satellite into Earth's orbit. In the case of DigitalGlobe this is around $150 million. The cost of the TerraBella's SkySats is around 50 million per unit. That is almost one tenth of the cost of DigitalGlobe's satellite and one hundred of the cost of a KH-11.

The launching costs are significantly lower. TerraBella, on September 15, 2016 launched four satellites in the same rocket. As the recent launch of the Indian Polar Satellite Launch Vehicle, PSLV carrying 104 satellites (including Planet's 88 Doves, BBC News [4]) shows, the market is changing fast enough that relatively inexpensive launches should be available, which in turn permits the testing of newer technology in each step.

Conclusion

Space assets are an essential part of any Intelligence, Surveillance and Reconnaissance capabilities. They provide unrestricted and accurate capabilities which multiply the effectiveness of any civilian or military force involved in the security and defence domain.

Through the years space assets have had so far a growing involvement in ISR functions. As space systems have a "dual use" character, this chapter focused on the commercial Earth Observation satellites, arguing that this involvement can be parsed in three distinctive generations.

The first generation started in 1972. Commercial EO took most people by surprise. This generation lasted longer and went further than anyone expected. After an extended 26 year period, new systems with very high resolution appeared for the first time in 1999 putting the first generation systems on the sidelines. The second generation, which started with the launching of Ikonos, saw the maturation of spatial resolution and geolocation accuracy of its imagery and lasted 14 years. But high intensity operations and full tactical needs remained hardly satisfied by second generation space systems yet, few believed at that time that commercial EO satellites could get any better in terms of temporal or spectral resolution.

Nowadays, we may be seeing the start of a new generation of multiple smaller and cheaper satellites than the existing ones, which can give imagery in almost real time. This is the result of miniaturization and use of high technology. Miniaturization of electronics is a mega-trend which will make sensors smaller and smaller, less costly, smaller, lighter and less power-hungry.

In our judgement, the third generation began in 2014. The Goliath of the space industry currently, DigitalGobe, has already reached twice the world's highest-resolution commercial earth observation satellite with a ground resolution of 0.31 m in panchromatic mode. This is the sharpest limit the US government permits for commercial satellite imagery. Although it is still the centerpiece of the GEOINT business in commercial EO imagery, it is obvious that the Davids of the industry, the emerging players with small satellites constellations which have high revisit-temporal resolution and videos too, are preparing to significantly contribute to it (Fig. 8.8).

This third generation is still in its early stages, it is almost in stealth mode and going on. Yet, few outside the U.S., are aware of this. The situation is still fluid and it reminds the one back in 1994. Then various companies such as Space Imaging, Orbital Sciences, GDE, Earth Watch and Resource 21 were pioneers in the high-resolution remote sensing satellites domain [14]. Lot of things has changed since then. Consolidation prevailed and only few of them survived.

There are already some similar signs. In February 2017 Google's parent company Alphabet agreed to sell its TerraBella satellite business to Planet [22]. Google is taking a stake in the Planet, while it has also agreed to purchase satellite images from it in a multiyear deal. Also the same month Canadian satellite company MDA bought US-based DigitalGlobe [10, 29]. So it is safe to assume that competition will

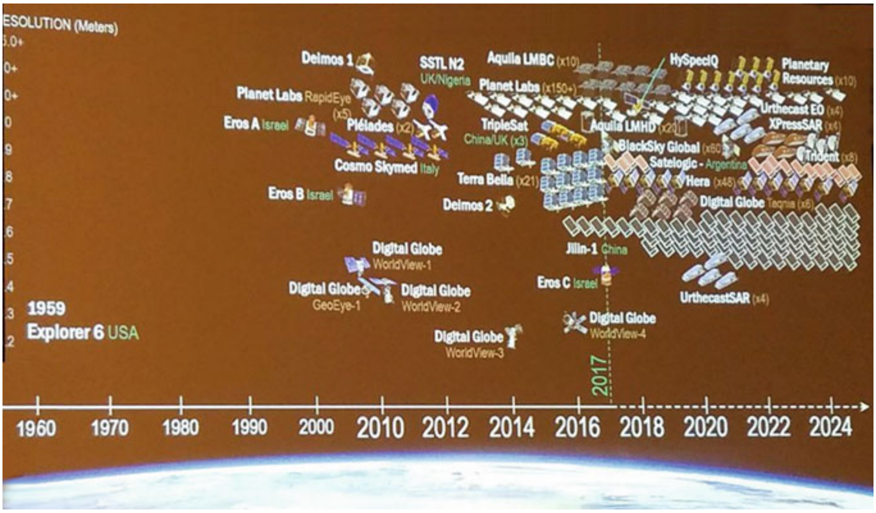


Fig. 8.8 Image taken from a NGA's deputy director Sue Gordon speech given at SXSW 2017: A space odyssey. Changing our view of earth [15]

intensify. It is reasonable to expect that the EO market while it will have new products and services which will positively impact the growth of this sector, not all companies will survive.

But whatever happens over the next decades, we may be able to look back and identify 2014, as an important turning point. If so, who knows how long this one lasts? According to Moore's Law, the golden rule for the electronics industry, the computing power tends to approximately double every two years. Although the improvement in optics do not keep the same pace with the electronics, if we judge from the past trend, this third generation will be shorter than the previous ones, too.

The technology improvement actually breeds new technology on its own. But the evolution towards increasingly higher commercial resolutions (in terms of spatial, temporal and spectral) seems to be a continuing quest. This is not just about the proliferation of satellites. Obviously, multiplying the potential sources of satellite EO data would also considerably improve the data latency needed in periods of crisis. This is about new advanced capabilities bringing a revolution in geospatial information with profound implications to users in the security and defence domain.

Their output will be massive datasets of global coverage, updated initially weekly, when compared to more traditional means of collection used in the past or present. In the past, such big geospatial data could not be easily handled by the existing database management tools, as they require strong computing power and intelligent software to analyze them and look for trends and valuable pieces of information.

This is a big data and data analytics play which presumably will change information needs and demands. This is about artificial intelligence/machine learning and its implications in intelligence. It is our belief that this revolution in the expanding capability and affordability of multisource information from commercial space-based earth observation systems will grow exponentially. And as such it will change the face of intelligence. This in turn will assist better decision making to meet emerging security and defence challenges.

References

1. American society of photogrammetry (1982) Manual of remote sensing, 2nd edn, vol. I, pp 20–24
2. Application of surveillance tools to border surveillance ‘concept of operations’. 7 July 2011. European commission—enterprise and industry 8 January 2012. https://ec.europa.eu/research/participants/portal/doc/call/fp7/fp7-space-2012-1/31341-2011_concept_of_operations_for_the_common_application_of_surveillance_tools_in_the_context_of_eurosur_en.pdf
3. Baker JC, O’Connell KM, Williamson R (2001) Commercial observation satellites: at the leading edge of global transparency. Santa Monica, CA: RAND Corporation. http://www.rand.org/pubs/monograph_reports/MR1229.html
4. BBC News, India launches record 104 satellites in single mission, 15 February 2017. http://www.bbc.com/news/world-asia-india-38977803. Accessed 20 February 2017
5. Book E (2003) Non-US firms provide niche imagery products, May 2003. http://www.nationaldefensemagazine.org/archive/2003/May/Pages/Non-US3877.aspx. Accessed 20 February 2017
6. Bowles D (2016) The facts about UrtheCast’s democratization of earth observation: who we are—and aren’t. https://blog.urthecast.com/updates/facts-urthecasts-democratization-earth-observation-arent/. Accessed 20 February 2017
7. Brown G (1987) International cooperation in space enhancing the world’s common security. Space Policy 174
8. Brown J (2015) The greatest view of earth: one company’s stranglehold on satellite imagery, exponential tech investor, bonner and partners. https://bonnerandpartners.com. p 8 November 2015. Accessed 2 June 2016
9. DigitalGlobe1 (2017) Inc, Press Releases, June 5, 2017. http://investor.digitalglobe.com/phoenix.zhtml?c=70788&p=irol-newsArticle&ID=2278719. Accessed 20 July 2017
10. Dow J (2017) Canada’s MacDonald Dettwiler to buy DigitalGlobe, February 17, 2017. http://www.reuters.com/article/us-digitalglobe-m-a-macdonald-dettwiler-idUSKBN15W2AJ. Accessed 20 February 2017
11. Farley G (2017) BlackSky releases first satellite Earth images, king, December 27, 2016. http://www.king5.com/tech/blacksky-releases-first-satellite-earth-images/379446035 Accessed 20 February 2017
12. Florini AM (1988) The opening skies: third-party imaging satellites and US security. Int Secur 13(2):100
13. Foust J (2015) UrtheCast releases high-definition video from space station camera, June 17, 2015. http://spacenews.com/urthecast-releases-high-definition-video-from-spacestation-camera/#sthash. Accessed 20 February 2017
14. Fritz LW Commercial earth observation satellites. In: paper presented at the XVHIth congress, Vienna, Austria, July 9 to 19, 1996. http://www.isprs.org/proceedings/XXXI/congress/part4/273_XXXI-part4.pdf. Accessed 20 February 2017

15. Gordon S (2017) SXSW 2017: A space Odyssey. Changing our view of earth conference. Transmitted live from NGA on the web. https://www.facebook.com/NatlGEOINTAgency/videos/1416486948395660/?comment_tracking=%7B%22tn%22%3A%22O%22%7D. Photo taken from tweet of KM Connifey @kaminskikdg 15 March 2017, retweeted by the official Twitter account for the National Geospatial-Intelligence Agency.com. Accessed 15 March 2017
16. Grundhauser LK (1998) Sentinels rising. Commercial high-resolution satellite imagery and its implications for us national, Security Air Univ Maxwell Afb Al Airpower Journal
17. Ha A (2017) satellogic aims to launch a constellation of small imaging satellites around Earth, June 20, 2014. https://techcrunch.com/2014/06/20/satellogic-launch/. Accessed 10 Jan 2017
18. DigitalGlobe2 (2017) http://explore.digitalglobe.com/see-freedom.html?mkt_tok=eyJpIjoiWVdVME1EaGlOVEU1Tm1RMiIsInQiOiI0aTFSa3ZjNEI0NjJiWmxYWW5sYktFS1RSMHViTHFraVJIZUhWMzBBa2JMNFVPd1l2c3FFd3lDVHZzWkJUN3pWRFk5TFwvVVBoWU5xVENDTlZXWTMwaUtHYnJcL2VVakh6RWpxOGFiUFBhWDRtTGdtSlwvSTk0dW5paTZpMHRsSVRBNCJ9. Accessed 7 February 2017
19. Kiran M, Shearn M, Smiley B D, Chau AH, Levine J, Robinson D (2014) SkySat-1: very high resolution imagery from a small satellite, Proc. SPIE 9241, sensors, systems, and next-generation satellites XVIII, 92411E (October 7, 2014). https://doi.org/10.1117/12.2074163. http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1916178. Accessed 20 February 2017
20. Kolovos A (1992) Gulf war: a critical evaluation of the role of space systems. hellenic national center for space applications, occasional study no. 3 (in Greek). https://haf.academia.edu/AlexandrosKolovos
21. Lee WT (1994) Unlocking the secrets of Krasnoyarsk, The Washington Times, 8/2/94
22. Nicas J (2017) Google to Sell Satellite Unit to Planet Labs, February 3, 2017. https://www.wsj.com/articles/google-sells-satellite-unit-to-planet-labs-1486156850. Accessed 20 Feb 2017
23. Peter N (2009) European space activities in the global context. In: Schrogl KU, Mathieu C (eds) Yearbook on space policy 2007/2008: from policies to programmes. European Space Policy Institute, Springer, Wien, NewYork, pp 2–114
24. Richelson J (2003) The satellite gap, bulletin of the atomic scientists, sage, 59: 4. http://journals.sagepub.com/doi/pdf/10.2968/059001014
25. Rickards J (2016) Satellite wars: the making of the next defense boom, strategic intelligence, March 2016 https://agorafinancial.com. Accessed 20 March 2016
26. Safyan M (2017) Planet to launch record-breaking 88 satellites, February 3, 2017. https://www.planet.com/pulse/record-breaking-88-satellites/. Accessed 20 Feb 2017
27. Sandau R (2011) Implications of new trends in small satellite development. In: Schrogl KU, Pagkratis S, Baranes B (eds) Yearbook on space policy 2009/2010: space for society. European Space Policy Institute, Springer, Wien, NewYork, pp 296–312
28. Steinberg GM (1998) Dual use aspects of commercial high-resolution imaging satellites. Begin-Sadat center for strategic studies
29. Vynck G, Deveau St, MacDonald Dettwiler Buying DigitalGlobe for $2.4 Billion, Feb 24, 2017. https://www.bloomberg.com/news/articles/2017-02-24/canada-s-macdonald-dettwiler-buys-digitalglobe-for-2-4-billion, Accessed 20 Feb 2017
30. Williamson RA, Baker JC (2004) Current US remote sensing policies: opportunities and challenges. Space Policy 20:109–116
31. 1988 SPOTIMAGE Corporation, SPOT: Acquisition Flexibility + Revisit Flexibility + Stereo Capability + High Spatial Resolution = Surveillance, (Toulouse: SPOTIMAGE Corporation)

Chapter 9
Drones Surveillance—Challenges and Techniques

Anastasios Kokkalis and Theodore I. Lekas

Abstract The increasingly complex asymmetric threats against people or infrastructure demand continuously improving surveillance strategies. Such strategies should be based on past experience, available technological capacities and foreseeable technological improvements and breakthroughs. Since rarely can either the location or the time of an attack be anticipated, efficient real-time surveillance of a potential target is of paramount importance. Such surveillance is usually performed by dedicated devices carried on manned or unmanned land, sea or air platforms. The use of manned platforms for such tasks could entail high risks for the crew, while unmanned ones can operate with considerable less constrains and in harsher environments. This later type of platforms are in fact "robots" able to be either remotely controlled or operating autonomously. The ensuing chapters present the key technological and operational elements that make remotely controlled and/or autonomous vehicles ideal for the execution of surveillance tasks for both civil and military applications and the challenges, be it technical or ethical, that such operational applications may face.

Introduction

Surveillance is defined as the close observation of a person, a group of persons or of a location in order to gather information for defensive or offensive purpose(s) using dedicated devices. The above can be carried out by humans or by manned or unmanned platforms which can be land, sea or air vehicles.

In particular, surveillance by air can be carried out either by means of manned platforms, such as airplanes, airships or helicopters, or by unmanned ones, such as satellites, UAVs, blimps or drones. The later are components of what is called UAS

A. Kokkalis (✉) · T.I. Lekas
Hellenic Air Force Academy, Dekelia Air Base, Dekelia, Attica, Greece
e-mail: tasos.kokkalis@hafa.haf.gr

T.I. Lekas
e-mail: theodoros.lekas@hafa.haf.gr

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_9

(Unmanned Aerial Systems). Manned airplanes and helicopters are big energy consumers and are difficult to remain unobserved, while the emotional status of the crew can compromise the outcome of a mission. Satellites are an expensive solution and in cases where observation inside a building is required, for instance when hostages and/or terrorists are located inside a building, are totally ineffective. In such situations, drones are very efficient platforms if correctly selected for the specified mission. In the ensuing chapters, we will concern exclusively on the historic development and use of drones for the execution of observational tasks for both civil and military applications.

Drones can be classified as flying robots. More precisely, they are devices guided either by remote control via a trained operator or by onboard computers. The guidance mode can be preset before launch or switched in flight. Drones utilized a host of new technologies such as digital radio control, wide-bandwidth wireless communications, miniaturized engines (electrical or IC), inertial navigation technologies based on laser gyroscopes, specialized electronic sensor devices and advanced materials are some of the common technologies involved. All these technologies combined, dictate the shape, the size and the performance of a drone designed for a specific mission. Drones can be heavier or lighter than air. The former can be with fixed, rotary or flapping wings.

There are no universally accepted criteria for Drone classification. Instead, each defense or civil organization creates its own criteria. Generally speaking, Drones (or UAVs or UASs) may be classified by size, range and endurance. The military uses a specific classification system based on range and/or endurance. Specific examples of the various classification systems used are provided below.

A classification by size comprises of the following sub-categories (Dalamagkidis [1]:

- Very small UAVs (insect size to 30–50 cm long) and which can be sub-divided as
  - Micro or Nano UAVs (up to 15 cm long)
  - Mini UAVs (between 15 and 50 cm long)
- Small UAVs (at least one dimension from 50 cm to 2 m, payload around 5 kg)
- Medium UAVs (wingspan 5–10 m, payloads 100–200 kg)
- Large UAVs (Up to now they are mainly used for military operations)

If classified by range or endurance, the following categories are generally accepted (classification per the US military):

- Close range, short endurance UAVs (range of 5 km, endurance 20–45 min)
- Close range, medium–to–long endurance UAVs (range of 50 km, endurance of 1–6 h)
- Short Range UAVs (range of 150 km, endurance 8–12 h)
- Mid-range UAVs (range of 650 km)
- Endurance UAVs (range of 300 km, endurance of 36 h)

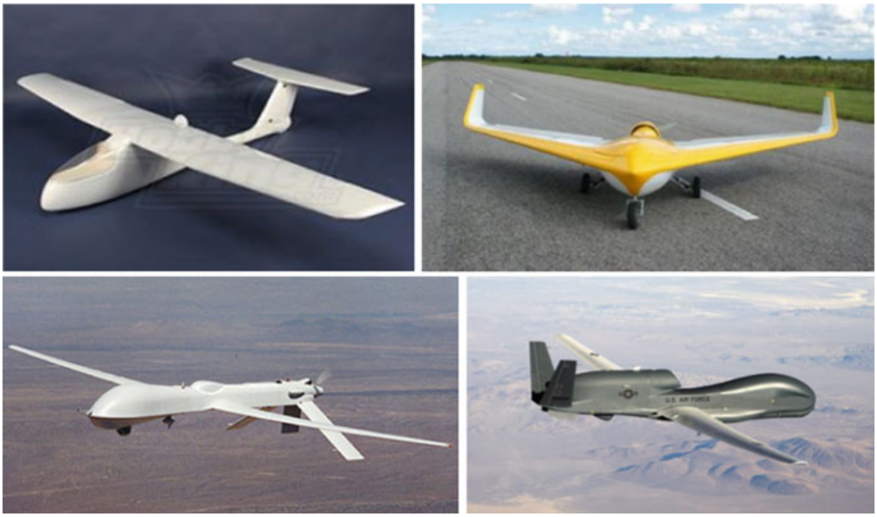


Fig. 9.1 Fixed wing drones

Examples of modern drones of various sizes, configurations and capabilities are presented in Figs. 9.1, 9.2, 9.3, and 9.4 (taken from the Internet).

The use of Drones as instruments of war has a long history going back to the mid of 3rd century AD with the use of the “Kongming lanterns” by the Chinese army for military signaling and for frightening their enemy away from the battlefield. In

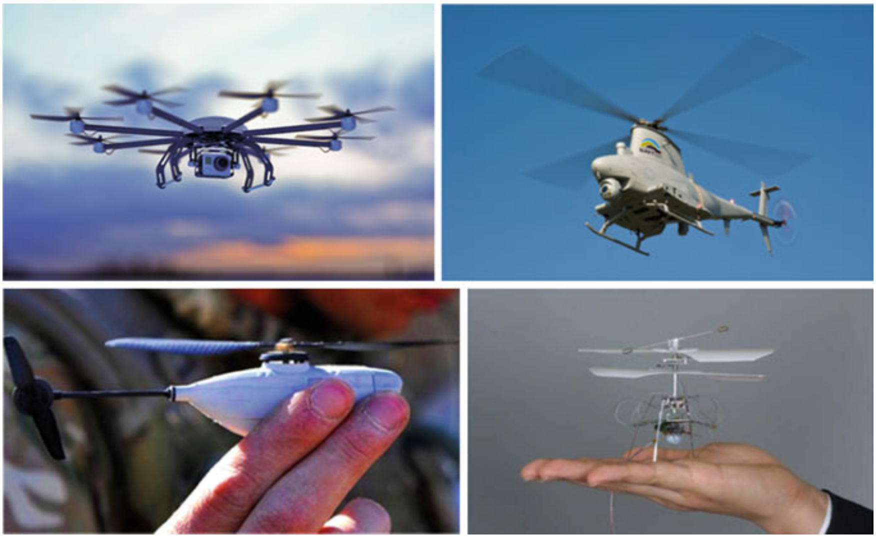


Fig. 9.2 Rotary wing drones

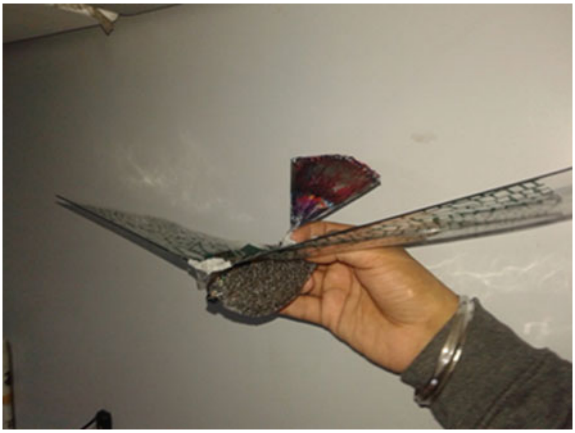


Fig. 9.3 Flapping wing drone

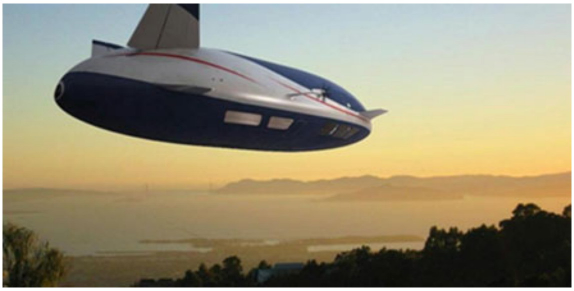


Fig. 9.4 Lighter than air drone

Europe, it can be said that the first act of war using drones took place in 1849, when Austrians sent unmanned balloons full of bombs to attack the city of Venice. During World War I, the US Dayton—Wright Airplane Company designed and built a torpedo launched by an airplane and that could explode at a preset time. Combat aerial unmanned vehicles were described by Nicola Tesla in 1915, but the first attempt to build a powered unmanned aerial platform was an aerial target in 1916. The fast-technological advances that ensued after World Wars I and II, permitted the development and building of more and more sophisticated and operationally efficient drones of various sizes, configurations, capacities and propulsion modes that have seen an ever-increasing use both in the Civil and Military fields of application.

As surveillance platforms, drones can be used for law enforcement, for monitoring actions in hostage-holding situations or for specific civil or military missions. Some of them, having small dimensions, can fly in narrow and confined spaces while producing minimal noise and presenting a target signature which is very difficult to detect. Typical civilian missions can be close monitoring of illegal acts, border surveillance, road traffic monitoring/control, environmental monitoring for air or sea pollution, wild life monitoring, agricultural monitoring, woodland fire prevention, building energy efficiency monitoring and monitoring of a devastated area after the occurrence of a natural disaster. Missions' specific to military interest are electronic signal intelligence, battle field monitoring, battle damages assessment and monitoring of sensitive areas.

Modern drones use a host of advance lethal and non-lethal sensors and/or devices to carry out the above-mentioned missions. Examples of both non-lethal and lethal devices are Wi-Fi crackers, fake cell phone towers, signal jammers, infrared cameras, live—feed videos, heat sensors, radars, Tasers, rubber bullets guns and lethal offensive devices like air-to-ground missiles. Due to some of the above-mentioned capabilities, the use of drones could raise significant ethical and legal issues with regards to privacy and civil liberties in addition to military code-of-contact. In many instances, it is very difficult if not impossible to make the distinction between military and civil use of a drone. For instance, there is absolutely no difference in the use of a Drone for the tracking of a fast boat involved in an illegal trafficking action and the tracking of a fast boat carrying enemy Special Forces personnel.

Acquisition and Operational Cost Versus Capabilities

The emphasis that may be placed between acquisition and operational costs versus the system's operational capabilities depends very much on the intended use of the Drone system, i.e. if it is designed primarily for civil or military use. For instance, a civil surveillance drone system, to be operationally attractive, it should have the lowest possible acquisition and operational cost combined with a high operational versatility, reliability and survivability. In contrast, for a military surveillance drone system to be attractive to its operator(s) it should be of rugged construction, offer new capabilities in the area of operational environment (e.g. day/night/smoke etc.) and ease of operational control (e.g. short period of operator training, single user operation etc.).

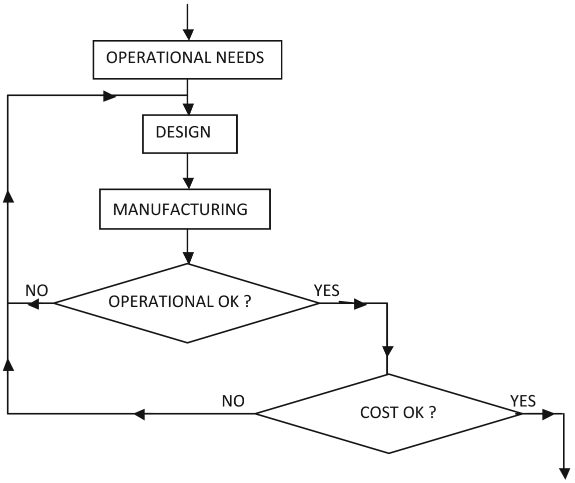
Irrespective of civil or military field of use, however, all the above requirements/characteristics depend primarily upon the following three parameters:

- accurate operational needs assessment,
- appropriateness of design (i.e. fit-for-purpose) and,
- non-complicated manufacturing and assembly methods.

These parameters are inter-dependent and interact with each other during the design stage of a Drone until convergence to an optimum solution is obtained, i.e. one that combines the highest operational capabilities with the lowest acquisition and operational cost [7]. The above-described interaction is shown schematically in Fig. 9.5.

An assessment of the operational needs must be done at first to accurately determine all vehicle and systems design constrains i.e. payload weight and

Fig. 9.5 Iterations to an optimum solution



dimensions, vehicle size, flight speed and altitude, range, endurance and onboard systems power requirements. This assessment must be based on a multiplicity of user requirements such as those gathered from field experience of the potential user(s), anticipated future operational needs and the personnel experience of the design team. This means that drones cannot be developed independently of the potential users' expectations and the design team's experience of a whole security strategy.

Further, a successful design should include not only the state of the art on existing devices for specific mission profiles but should also sustain sub systems upgrades through minimum, if any, modifications of the initial vehicle frame Unmanned vehicles, handbook [4–6, 10]. The required operational versatility dictates that the onboard systems must be multifunctional and if not, their replacement per specific mission requirements should be fast and straightforward. In this way the need for acquisition of a new drone system is avoided. Only a new sub system would be acquired, thus lowering the upgrade acquisition costs.

In addition a successful design should ease field repair tasks by facilitating the quick replacement of deficient component(s). This implies the modular internal and external design of the vehicle and its payload. Vehicle and system modularization will ease the burden of maintenance as well as make the training for maintenance personnel less demanding. Thus, the training time and costs will be significantly reduced since the need for specialist maintenance personnel will be minimized if not eliminated altogether, leading to significant reductions in the entire operational life-cycle costs. Furthermore, modularization combined with ease of maintenance implies the need for "simple designs" which in turn will permit the use of simple manufacturing processes, thus leading to lower production costs. In addition, modularization of the design will shorten the vehicle assembly and disassembly time as well as decrease the number of spare parts needed to support an operation.

A drone should be operable either by remote control or autonomously within a wide range of environmental and operational conditions and with a high level of survivability built in the system. This dictates the use of a robust, light weight and efficient Flight Control System (FCS). The required robustness and light weight are achieved through a fail-safe design, miniaturization of the constituent components and the use of advanced materials, either composite or light alloys. The required efficient control may be achieved by means of leading edge aerodynamic and structural design, fast running algorithms and quick responding flight control devices. The combination of the above will make a drone easy to handle on the ground and in the air without the necessity of lengthy training for either the ground support staff or the pilots. This will lead to further reductions in the overall operational costs. In addition, easy handling and maintenance could shorten the time needed to train and field operational support personnel as well as bring a complete drone system into the field of operations.

The higher the range and the endurance of the drone, the higher its operational versatility. One way of achieving this is by the use of an energetically efficient and reliable propulsive system. For the small size Drones the power comes usually frombatteries (e.g. LiPo or Li-ion) whilst for the medium and large size Drones can come either from a Hybrid system (i.e. liquid fuel powered engine-electric power generator plus batteries for back up) or exclusively from a liquid fuel powered engine. For operational reasons, the propulsive system for a drone operating within the build environment or close to the ground should be as silent as possible. In case of a drone of very small size, some of its components should be multifunctional and with as much fault tolerance build-in as possible. For instance, flapping wings provide at the same time lift and thrust.

Typical Security Mission Requirements

The security challenges in a low intensity field of battle today are characterized by evolving asymmetric multi—theater terrorist threats as the experience, for instance in Afghanistan and Iraq, has shown. These threats necessitate a very agile and non-dogmatic response evolving the highest degree of technological approach. As it was stated earlier in this chapter, an unmanned surveillance air vehicle should be developed not in isolation but as a part of an integrated security strategy. The efficiency of such strategy mainly depends on the proper mix of the evolved means.

The primary mission of a drone, or a swarm of drones, is to provide to the commanding security officer near-real-time data on threat position, composition, and state of readiness and sometimes to proceed and take lethal or non-lethal actions. The exploitation of the results from a mission can be accomplished at various command levels. This exploitation depends on the onboard mission systems characteristics. An important part of a mission consists in gathering intelligence, which then should be exploited by the operation center. Drones, combined to other means like microsatellites [3], are now available for various support roles and are well suited for tactical, asymmetric operations, necessitating a simpler command and control network.

Drones are the preferred means of support for such operations because they require relatively fewer maintenance, control, and operating personnel or transportation assets for deployment in comparison to their manned counterparts. Their operational capabilities include day or night flight even under adverse weather conditions, survivability from hostile fire and secure radio links. They thus are able to provide capability for day or night intelligence, reconnaissance, surveillance and target acquisition (RSTA), rapid operation damage assessment (BDA) and near-real-time information. The gathered information may include video, freeze-frame, voice communications, recorded message traffic, and digital data. They also provide operational field management in high-threat zones where manned means would be exposed to high degree of danger because of possible hostile action. In case of lethal action been taken by a drone, it must provide for fire support at much greater distances than using forward observers or snipers so as to reduce the risk of such assets. Finally, Drones must have the capability to immediately participate to any part of the operation, not anticipated due to a rapid change in the tactical situation. From all the above statements, it can be concluded that a proper choice of a drone should be done according to the requirements of a specific mission.

Apart of the required technical merits/capabilities of the drone itself, the success of a drone operation depends also on several other parameters. The first one is the number of available systems. As the experience in Afghanistan and Iraq has shown (UAVs road map [9]), there must be enough systems available to satisfy the demand for simultaneous missions during the same operation. The lack of available systems can create conflicting priorities, thus compromising the success of a specific part of the operation or, even, of the entire operation. The second one is the number of frequency channels available. Very often, several drones should participate simultaneously in the same operation, at different locations and with different roles. The lack of enough free frequency channels can seriously compromise the near real time data transmission, thus possibly threatening the success of the entire operation. The third one are the meteorological conditions. High wind speeds and/or turbulence or high concentration of atmospheric particles such as rain droplets, ice crystals or dust can seriously compromise the mission by reducing the flight capabilities of the drone and/or the efficiency of its sensors.

Since drones can be components of an integrated security system, a centralized command and control system is required to ensure functional integration and mission prioritization. The efficient exploitation of the gathered information requires a comprehensive and integrated dissemination architecture and an optimized bandwidth usage. This way the satisfaction of the requirements is maximized. To provide situational awareness to all participants a net-centric approach to unmanned airborne system integration/interoperability is needed with a provision for archiving and discovery of full motion video collected. As it is obvious, situational awareness can be totally useless if not continuously refreshed.

Experience also has shown that in case of urban combat, high bandwidth wireless data communications are not suitable and that potentially loss of connectivity can result, even at short distances. This adverse effect is compounded by short Line-Of-Sight (LOS) distances, making visual reconnaissance difficult. It must be pointed out here, that urban environment is an extremely hazardous combat field due to narrow spaces and a lot of hiding places for the enemy. For this reason, the development of small, low altitude and autonomous unmanned air vehicles combined to the ability to coordinate their actions will be a useful tool in urban combat field.

For safety reasons, for instance in case of an operation nearby an airfield, the integration of the unmanned aviation into the national airspace is mandatory. This can be done through a close coordination between the national security and the civil aviation authorities. Such integration will also improve the responsiveness and the training.

An Example of Mission

As an example, the following surveillance mission scenario is proposed, concerning the monitoring of a sensitive area, such as an extended military base. It must be pointed out that such a network can also be set in an internal space, such as a huge hangar.

A common surveillance method includes a camera network, where cameras are placed at strategically selected positions. These cameras are static, so they can be targets to be destroyed by an intruder. If instead they could be mobile, they also would be harder to destroy. This can be achievable using MAVs carrying micro cameras, either optical or IR. One could imagine flight paths forming a network over the area to be covered. A swarm of MAVs are flying continuously along those flight paths in such a way that every node of this network will be surveyed at any moment by MAVs coming from a different direction.

The randomness of the direction of the incoming MAV will add to the difficulty of the motion of the intruder. Since MAVs have an autonomy restricted by their batteries capacity the avoidance of frequent takeoffs and landings for energy refueling should be addressed. The proposed solution, which is compatible to the actual technological capabilities, is the use of wireless power supply through laser beams [2, 8]. Since the MAVs will be assigned to fly at very low altitude, less than say 50 m, the required laser power can be easily affordable. In the MAVs design a laser beam receptor should be incorporated, from which the power supplied will be transmitted to the onboard batteries. Each time the stored energy will be lower than a prescribed threshold, the MAVs will fly over a specific point of the network where a laser device will be located.

References

1. Dalamagkidis K (2014) Classification of UAVs. Springer
2. Kawashima N, Takeda K, Matsuoka H, Fujii Y, Yamamoto M (2005) Laser energy transmission for a wireless energy supply to robots. In: 22nd international symposium on automation and robotics in construction, ISARC 2005-September 11–14, Ferrara (Italy)
3. Kolodziejski P, Microsatellites J (2015) UAVs: finding the right force mix. Booz Allen Hamilton microspace center of excellence Ph. 719–387-2029 May 6, 2015
4. Landolfo G (2008) Aerodynamic and structural design of a small nonplanar wing UAV, Master Thesis, University of Dyton, May 2008
5. Lissaman P (2009) Effects of turbulence on bank upsets of small flight vehicles. In: 47th AIAA aerospace sciences meeting including the new horizons forum and aerospace exposition, 5–8 January 2009, Orlando, Florida
6. Macheret J, Teichman J, Kraig R (2011) Conceptual design of low—signature high endurance hybrid—electric UAV, Institute for Defense Analyses, November 2011
7. Moumen I et al (2015) Design, fabrication and testing of a surveillance UAV. Adv Mater Res 1115:450–453
8. Nugent TJ, Kare JT (2010) Laser power for UAVs, white paper, laser motive L.I.C
9. Unmanned air vehicles road map 2005–2030, office of the secretary of defence
10. Unmanned vehicles, handbook 2008

Chapter 10
Early Warning Against Stealth Aircraft, Missiles and Unmanned Aerial Vehicles

Konstantinos C. Zikidis

Abstract Since the 2nd World War and during the Cold War, the air defense radar has proven to be the main surveillance sensor, where each radar would cover a radius of more than 200 nautical miles. Apart from the electronic warfare, more recently the emergence of stealth or low observable technology, the evolution of ballistic and cruise missiles, as well as the democratization of UAVs (Unmanned Air Vehicles) or drones, have contested the capabilities of the typical surveillance radar. All these targets are difficult to detect, because they exhibit low RCS (Radar Cross Section), potentially flying at the upper or lower limits of the radar coverage or outside the expected velocity range (being either too slow, e.g. some UAVs, or too fast, like ballistic missiles). This chapter begins with the estimation of the RCS of various potential targets, as a function of the radar frequency band. In this way, the expected detection range against a set of targets can be calculated, for any given radar. Secondly, different radar types are taken into consideration, such as low frequency band radars or passive/multistatic radars, examining the respective advantages and disadvantages. Finally, some issues are discussed concerning the "kill chain" against difficult-to-detect targets, in an effort to defend efficiently the air space.

Keywords Radar • RCS—radar cross section • Stealth • Low observable
Ballistic missiles • Cruise missiles • UAV—unmanned air vehicles • Drones
Low frequency band radar • Passive radar • Multistatic radar • Kill chain

Introduction

The invention of the radar system cannot be attributed to a single person or state. It is rather an evolution, pursued by many nations concurrently and sometimes antagonistically, before and after the 2nd World War. Development of radar

K.C. Zikidis (✉)
Department of Aeronautical Sciences, Hellenic Air Force Academy,
Dekelia Air Base, Greece
e-mail: kzikidis@cs.ntua.gr

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_10

systems culminated during the Cold War. The radar theory is now well established and a number of classic books are available, such as [1–3], while the radar has been considered as the primary sensor against aerial targets, for more than half a century.

Along with the radar, electronic warfare systems have also been evolving, exhibiting various features to counter radar operation [4, 5]. Today a typical jet fighter is equipped with a self protection system, including a Radar Warning Receiver (RWR), a multi channel jammer system, sometimes called ECM System (from the older term *Electronic Counter Measure*), as well as passive decoys (chaff and flare dispenser system) or even active towed decoys. Most jammers today include DRFM (Digital Radio Frequency Memory) capability, while more advanced counter-measure systems employ cross-eye jamming [6] or active cancellation techniques [7]. Of course, electronic warfare systems are not limited to aircraft but equip also ships, tanks or ground-based systems.

Since the late ′70s, a new technology appeared, gradually taking over the military world, even though it took more than a decade to come out: low observable or stealth. Following the "Have Blue" project, whose two prototypes proved the possibility to construct a stealth airplane (even though both crashed), the US began development of the F-117 Nighthawk and soon after the Advanced Technology Bomber (ATB) program, which eventually led to the stealth bomber B-2 Spirit. At the same time, they also decided to modify the B-1A to the B-1B Lancer, exchanging some of its high performance capabilities for the reduction of its radar signature. Since the end of the ′80 s, when the first stealth aircraft were revealed to the public, the reduction of the Radar Cross Section (RCS) has become the primary requirement for any military asset [8, 9].

Apart from jet fighters/bombers (either stealth or not), UAVs or Unmanned Aerial Vehicles, commonly (but not properly) called drones, have exhibited considerable advances during the last decades [10, 11]. Military UAVs are now amassing far more flight hours than manned fighter aircraft, while the Predator family UAVs (i.e., the MQ-1 Predator, MQ-9 Reaper and MQ-1C Gray Eagle) have accumulated until 2016 more than 4 million flight hours, according to their manufacturer. Such UAVs exhibit low RCS, while they may also fly too slow or too low, preventing some radars from detecting them. Even worse, the limits between a UAV and a loitering munition, e.g., the IAI Harpy and Harop, are rather obscure. Therefore, UAVs can be a quite hard-to-detect target, while armed UAVs (UCAV —Unmanned Combat Aerial Vehicle) and loitering munitions may pose a real threat to a radar system. On the other hand, a saturation attack by decoy UAVs, acting as jet fighters, would oblige fire control radars to switch on transmission (betraying their position and becoming vulnerable to anti-radiation missiles, as was the case in the 1991 Gulf War air campaign) or even trigger the launch of expensive surface-to-air missiles against low cost drones.

Cruise missiles can be considered as a similar threat for a radar, since they usually fly a few feet above the ground, but they feature better capabilities (speed, warhead, precision), while they also exhibit small RCS. On the other hand, ballistic missiles pose a substantially different threat, by flying very high in their apex phase and very fast during the final phase. Only a few radars have ballistic missile defense capabilities. Even in that case, combining speed with decoy deployment and maneuvering renders a ballistic missile difficult to intercept. In the case of a MIRV missile, featuring a payload of *Multiple Independently Targetable Reentry Vehicles*, it is rather impossible to avoid suffering damage.

All these technologies, i.e., electronic jamming, stealth, UAV/UCAV/etc. and cruise/ballistic missiles, challenge the capabilities of radar systems. In an effort to ensure the surveillance of a given air space, the following steps will be considered:

1. First, the RCS estimation for various targets is attempted. Apart from open source RCS values, an approach based on computer simulation will be followed for the F-16 and the F-35 jet fighters, as well as for the DF-15 short-range ballistic missile, as case studies. More specifically, a 3D model will be constructed for each of these targets. Then the POFACETS code, a MATLAB application based on the Physical Optics method, will be employed to predict the RCS of these targets, in certain radar frequency bands. This approach has been proposed in [12, 13].
2. Secondly, the expected detection range for a typical ground radar will be estimated against various threats. In this way, the significance of stealth technology will be proven.
3. Different radar types will be examined, such as low frequency and passive radars, complementing each other, in order to cover the air space.
4. The chapter will conclude with an analysis of the “kill chain” against stealth threats and how not to break it, taking into account also operational issues.

RCS Estimation of Various Targets

The RCS (Radar Cross Section) of a target can be defined as the projected area of a virtual metal sphere which would scatter the same radiation power as the target does [8]. The RCS is usually represented by the symbol σ and expressed in square meters (m^2) or in dBsm (decibel with respect to 1 square meter). It depends on the actual size, the shape and the reflectivity of the target (i.e., the coating). Taking into account the radar equation, the range at which a radar detects a target is proportional to the 4th root of the target’s RCS, as shown below.

The RCS exhibits significant fluctuations and can be considered as a stochastic function of the relevant position and aspect of the target with respect to the radar. A mean value can be computed for the front sector of a target, which could be used to calculate the maximum detection range of this “incoming” target for a certain radar set.

The RCS of any military asset is supposed to be classified. However, mean frontal RCS values for various targets have appeared in the open literature, either from measurements in suitable test ranges or from theoretical estimation with the

Table 10.1 RCS values for various targets. These values are just indicative, presumably referring to the frontal aspect (head on) RCS of an aircraft in clean configuration (without external loads, such as fuel tanks, missiles etc.), in the X-band (8–12 GHz) [9]

Target	RCS (m^2)
Navy cruiser (length 200 m)	14000
B-52 Stratofortress	100–125
C-130 Hercules	80
F-15 Eagle	10–25
Su-27 Flanker	10–15
F-4 Phantom	6–10
Mig-29 Fulcrum	3–5
F-16A	5
F-18C/D Hornet	1–3
Mirage 2000	1–2
F-16C (with reduced RCS)	1.2
T-38 Talon	1
B-1B Lancer	0.75–1
Sukhoi FGFA prototype (derivative of PAK FA for India)	0.5
Tomahawk TLAM	0.5
Exocet, Harpoon	0.1
Eurofighter Typhoon	0.1 class
F-18E/F Super Hornet	0.1 class
F-16IN Super Viper (proposed to India for the MMRCA)	0.1 class
Rafale	0.1 class
B-2 Spirit	0.1 or less
F-117A Nighthawk	0.025 or less
bird	0.01
F-35 Lightning II	0.0015–0.005
F-22 Raptor	0.0001–0.0005
insect	0.00001

help of computational electromagnetics, or even from unofficial leaks. In Table 10.1, a comprehensive RCS list for several targets is depicted, based on [9] and the references therein. It is noted that most of these values are given “as is”, since there is no official claim or statement.

Trying to examine the RCS for some representative targets and its dependence on the radar frequency, an approach based on computational electromagnetics was followed. The term “computational electromagnetics” refers typically to computationally efficient approximations to Maxwell’s equations in order to obtain real life results, since it is quite difficult to obtain closed form solutions in real world problems, unless the related physical objects are very simple.

Concerning the issue of RCS prediction of a physical object of known shape, there are various methods, such as the Method of Moments, the Finite Difference

Method, Geometrical Optics and Physical Optics. The last one, the Physical Optics (PO) method, yields good results at higher frequency bands (closer to the optical region), in the specular direction, by approximating the induced surface currents. The PO currents are integrated over the illuminated portions of the target to obtain the scattered far field, while setting the current to zero over the shadowed portions. Despite some certain shortcomings, the simplicity of the PO method ensures low computational overhead [14, 15].

The POFACETS 4.1 code is a MATLAB application, developed by the US Naval Postgraduate School, implementing the PO method for predicting the RCS of complex objects. The program models any arbitrary target by dividing it to many small triangular facets and the scattered field of each facet is computed as if it were isolated and other facets were not present. Multiple reflections, edge diffraction and surface waves are not taken into account. Shadowing is included by considering a facet to be completely illuminated or completely shadowed by the incident wave. The PO method is used to calculate the induced currents on each facet. The scattered field is computed using the radiation integrals [16].

Apparently, in order to employ the POFACETS code, it is necessary to have a 3D model of the target. Taking into account that reliable blueprints or CAD files cannot be available for any military asset, except maybe for very old ones, a 3D model of the target has to be created. The procedure used for that purpose is described in [12, 13] and in a few words is as follows:

- Preprocessing of 2D images or still images from videos of the object, by converting them to drawings, using software such as GIMP.
- Estimation of the overall dimensions of the object.
- Construction of a properly scaled 3D model, based upon the above-mentioned drawings and dimensions, with the use of suitable software, e.g., AUTODESK 3ds Max or Blender 3D suite.
- Fine-tuning of the 3D model, based on photos/videos.
- Running simulations with the POFACETS code, which imports .stl files and converts them to .m files, to be processed by MATLAB. The imported 3D models are considered to be *Perfect Electric Conductors*. The monostatic RCS is computed, where the transmitter and the receiver are co-located, as is the case for most radar systems.

RCS of the Lockheed Martin F-16C

Following the procedure analytically described in [13] and briefly mentioned above, the 3D model of an F-16C is depicted in Fig. 10.1. The aircraft is depicted with its radar nose cone but in the computer simulation the model used was without the cone, which is more or less transparent for the radar. The RCS of the F-16 model was computed for a radar transmitting at 10 GHz (X-band), like a typical fire

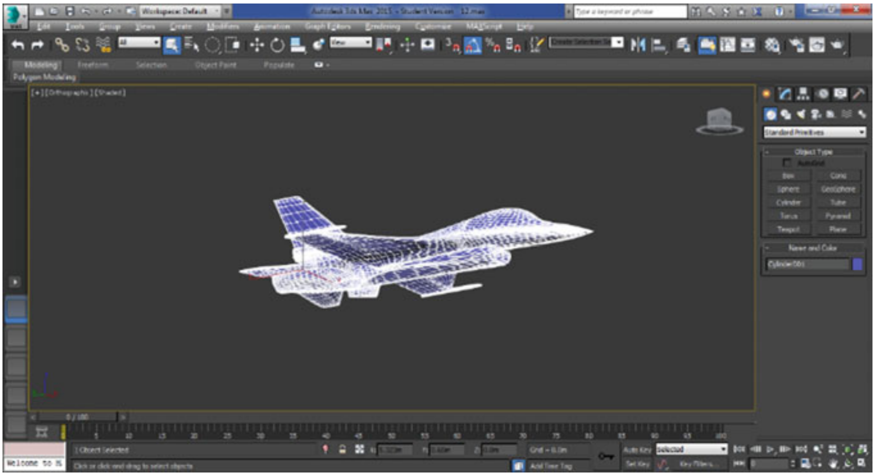


Fig. 10.1 The F-16C model, with the radar nose cone, created with the help of AUTODESK 3ds MAX software [13]

control radar. The RCS pattern shown in Fig. 10.2 corresponds to the polar plot of the F-16 RCS, seen at the same level (θ = 90° and φ ranging from 0° to 360°). The mean frontal RCS, averaged from −30° to +30° in azimuth (in steps of 1°) and from −15° to +15° in elevation (in steps of 5°) is −2.8 dBsm (that is 0.525 m^2).

According to Table 10.1, the mean RCS of the F-16C is reported to be 1.2 m^2, while the RCS of the F-16IN proposed in the frame of the recent MMRCA competition in India is in the 0.1 m^2 class, featuring an AESA radar and possibly other RCS reduction treatments as well. Therefore, the above-mentioned result (based on an F-16 model with AESA radar) is quite reasonable, falling between these two RCS values. If further RCS reduction measures had been taken into account, such as the application of the HAVE GLASS program, the RCS would be even smaller, approaching the reported F-16IN RCS value.

RCS of the Lockheed Martin F-35

For the F-35, a similar approach was followed and two models were created, one with and one without the radar nose cone [13]. The model with the nose cone is depicted in Fig. 10.3. However, the model without the nose cone was imported to the POFACETS code and the resulting RCS polar plot, seen from 10° below, is depicted in Fig. 10.4. It is noted that the use of radar-absorbent material (RAM) coating, which would further decrease the RCS, has not been taken into account.

The mean overall RCS and the mean front sector RCS (averaged from −30° to +30° in azimuth, in steps of 1°, and from −15° to +15° in elevation, in steps of 5°)

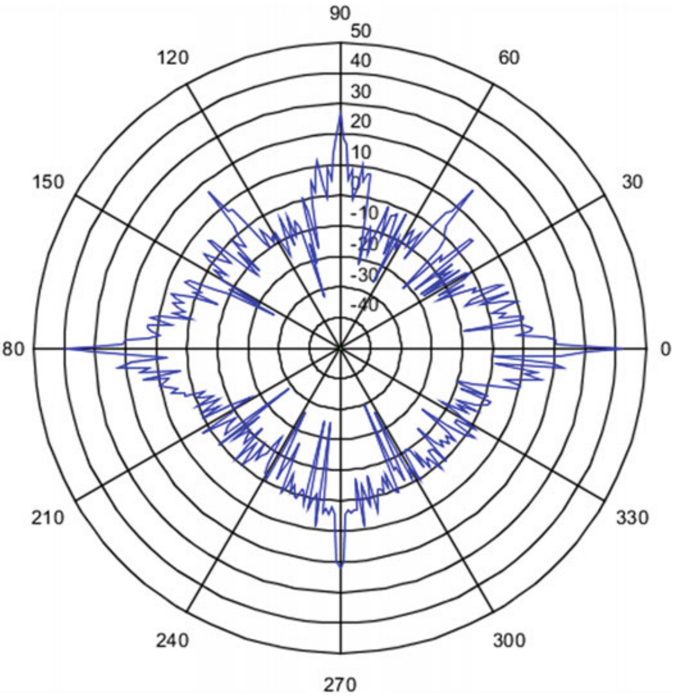


Fig. 10.2 RCS polar plot for the F-16C model, at the same level, at 10 GHz (the aircraft nose is pointing at 90°) [13]

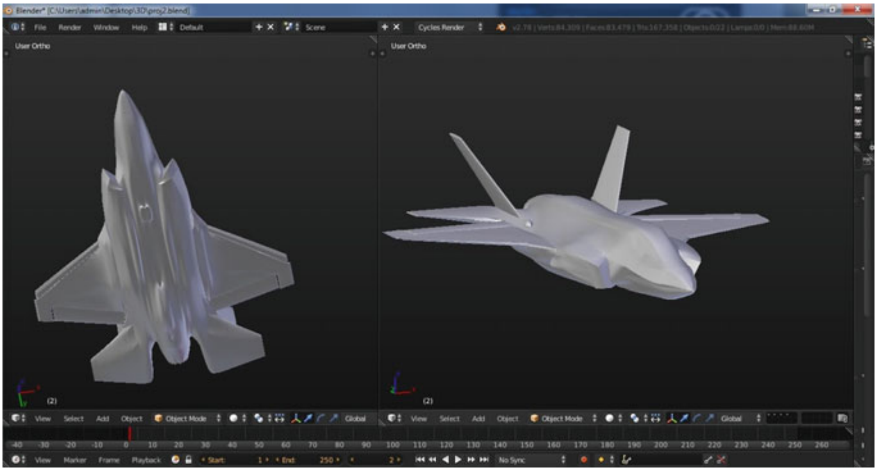


Fig. 10.3 The F-35A model with the radar nose cone, created with the help of Blender 3D suite [13]

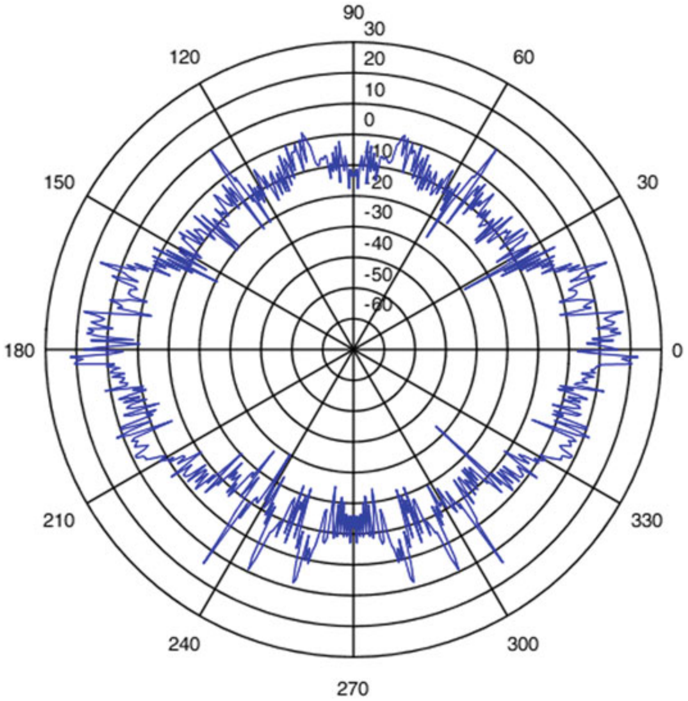


Fig. 10.4 RCS polar plot for the F-35A model, at 10 GHz, seen from below (depression angle 10°). The aircraft nose is pointing at 90°. RCS is relatively small in a wide sector in the front, apart from the peaks produced by the leading edges of the wings (approximately at 35° off-axis), attaining higher values at the sides (due to the wings and the fuselage) [13]

were calculated in various frequency bands, from VHF to Ku-band. The results are shown in Fig. 10.5. Obviously, the RCS of the F-35 is not so small at lower frequency bands.

The F-35 features advanced RAM (called “fiber mat” [17, 18]), which is more durable and requires less maintenance, with respect to coatings of older stealth aircraft, according to Lockheed Martin. The F-35 RAM has been reported to make use of carbon nanotubes (CNT) technology, absorbing electromagnetic waves over a wide range of frequencies [18]. Therefore, the actual RCS values are expected to be lower than the ones obtained by the POFACETS code.

RAM coatings are frequency selective, i.e., they provide higher attenuation at specific frequency bands, for example at the X-band or above. At lower frequency bands, RAM coatings are less effective. Trying to emulate the use of RAM, an attenuation in the class of 10 dB is considered, at least concerning the X-band and higher frequency bands.

As seen in Fig. 10.5, the mean RCS at 10 GHz (without the use of RAM) for the front sector is −10.1 dBsm. With the use of RAM, the RCS would be further

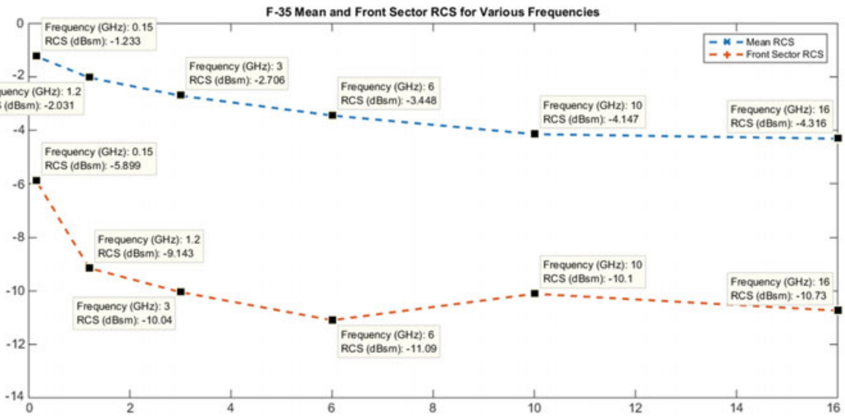


Fig. 10.5 The mean overall RCS (upper curve) and the mean front sector RCS (lower curve) of the F-35 model versus frequency. It is clear that the F-35 RCS is not so small at lower frequency bands. In this graph, the use of Radar Absorbent Material (RAM) is not taken into account. RAM would further reduce the RCS, especially at higher frequency bands (S-band and above) [13]

reduced to the class of −20 dBsm, which corresponds to 0.01 m^2, confirming that the F-35 exhibits a really small RCS. This value is higher than but quite close to (within 3 dB) the RCS values appearing in various sources, which estimate the front sector RCS of the F-35 from 0.0015 to 0.005 m^2 [9].

RCS of the Dong-Feng 15 (DF-15) Missile

The DF-15 is a Chinese short-range ballistic missile, in three variants (A, B and C). The 3D model of the DF-15C was created in CATIA v5 and is depicted in Fig. 10.6. Importing the model to POFACETS, the RCS diagram at 10 GHZ is shown in Fig. 10.7. Averaging the frontal RCS in a similar manner as before, the result at 10 GHz is at the class of −17 dBsm. By subtracting 10 dB, in order to emulate the use of RAM, the RCS becomes −27 dBsm, i.e., 0.002 m^2. At 150 MHz, the average head-on RCS reaches −13 dBsm (0.05 m^2). In the VHF-band, RAM is rather ineffective, without any significant RCS reduction.

The RCS of the DF-15 missile has been reported to be 0.002 m^2 in the X-band and 0.6 m^2 in the VHF-band [19]. In the X-band, the above mentioned result coincides with the reported RCS. At VHF, the computed RCS is higher than the one at 10 GHz but not exactly as the one reported in [19]. However, it is quite close, proving that the proposed approach yields reasonable results, as well as that the RCS of some stealth targets is considerably higher at lower frequency bands.

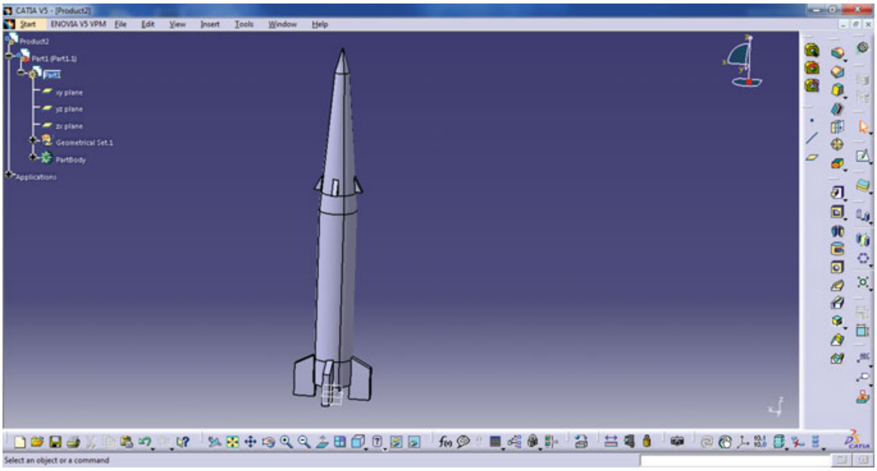


Fig. 10.6 The DF-15C model in CATIA v5 CAD suite [13]

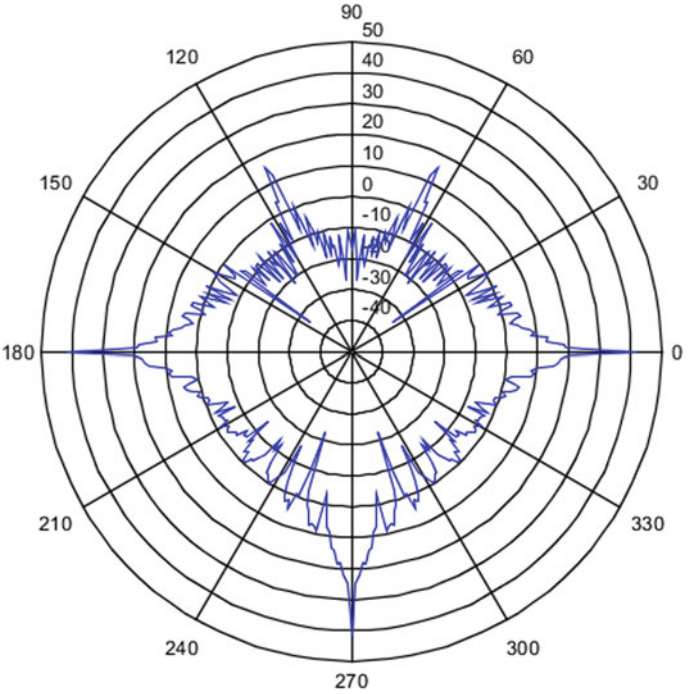


Fig. 10.7 RCS diagram for the DF-15C missile, at the same level, at 10 GHz (the missile tip is pointing at 90°) [13]

RCS Versus Detection Range

Considering the RCS values of Table 10.1, as well as the above-mentioned revized F-35 RCS values with the help of POFACETS, it is possible to make a diagram of the RCS for various targets. Furthermore, with the radar equation in mind, knowing the detection range of a radar for a certain target with known RCS, it is possible to calculate the respective detection range for any target.

More precisely, the fundamental form of the radar equation is as follows [1]:

$$R_{max} = \sqrt[4]{\frac{P_t G A_e \sigma}{(4\pi)^2 S_{min}}}$$

where R_{max} is the maximum detection range, P_t the transmission power, G and A_e the gain and the effective area of the transmitting and receiving antennae (which coincide in the usual monostatic radar), σ is the target RCS and S_{min} the minimum detectable signal. Therefore, for a given radar set, where P_t, G, A_e and S_{min} are fixed, R_{max} is proportional to the 4th root of σ: $R_{max} \propto \sqrt[4]{\sigma}$.

So, if a radar can detect a standard target with an RCS of 1 m^2 at $\hat{R}$, a target of σ m^2 will be detected at $R_{max} = \hat{R}\sqrt[4]{\sigma}$. For example, the Raytheon HR-3000 (HADR) S-band air defense radar can detect a 1 m^2 RCS target at 320 km or 173 nautical miles [9]. Assuming that the RCS values of Table 10.1 are valid for the S-band, a range vs RCS curve can be drawn, indicating also the various targets.

Instead of the "range vs RCS", an inverse "RCS vs range" curve is proposed, offering also a graphical representation of the detection range for each target. So, in Fig. 10.8 there is an RCS vs detection range curve for the HR-3000 radar. In this way, the range at which that radar can detect a target is shown in linear scale, starting from the left axis. Trying to depict the RCS of the various targets of Table 10.1, the gray rectangles shown in Fig. 10.8 are created, due to the uncertainty (min and max RCS estimated values, corresponding to different detection ranges).

It is noted that the HR-3000 operates in the S-band, while the values of Table 10.1 correspond to the X-band (a graph showing the IEEE radar bands is shown in Fig. 10.9). So, the RCS values in Fig. 10.8 are only indicative. Especially for the F-35, its RCS in the S-Band is approximately −10 dBsm or 0.1 m^2 (without RAM), according to Fig. 10.5. Taking into account that RAM is not as efficient at lower frequencies, an attenuation of 10 dB in the S-band would not be realistic. Assuming that the attenuation in the S-band is half as that in the X-band (i.e., 3 dB lower), the application of RAM would further reduce the RCS by 7 dB. Therefore, a more reasonable prediction of the F-35 RCS in the S-band would be −17 dBsm (that is 0.02 m^2). Such a target would be detected at more or less 65 nautical miles (n.m.) by an HR-3000 radar (and not at 40–50 n.m., as indicated in Fig. 10.8).

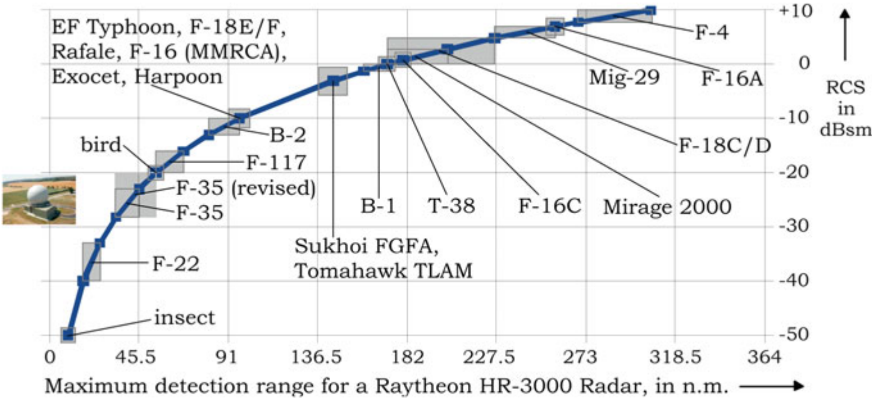


Fig. 10.8 RCS (in dBsm) of various targets versus the respective detection range (in nautical miles) for the HR-3000 S-band air defense radar. Each target is depicted with a gray rectangle, the size of which depends on the estimation uncertainty of its RCS. It should be noted that the RCS values of the targets are only indicative, since they correspond to a different radar band, i.e., the X-band [9]. In any case, the significance of low observable is obvious: while legacy fighters are picked up at more than 200 n.m. and modern jets (with RCS ≈ 0.1 m^2) at 100 n.m., stealth aircraft are detected at close ranges (< 65 n.m.)

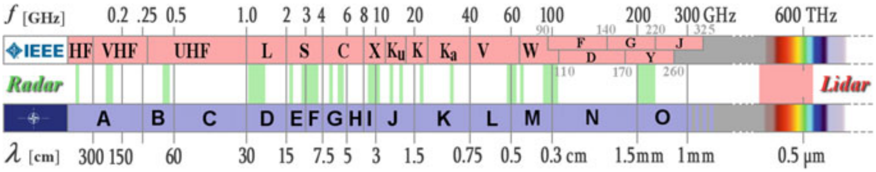

f [GHz]
0.2 .25 0.5 1.0 2 3 4 6 8 10 20 40 60 100 200 300 GHz 600 THz
90 140 220 325
IEEE
HF VHF UHF L S C X Ku K Ka V W F G J D Y
Radar
110 170 260
Lidar
A B C D E F G H I J K L M N O
λ [cm]
300 150 60 30 15 7.5 5 3 1.5 0.75 0.5 0.3 cm 1.5mm 1mm 0.5 μm

Fig. 10.9 IEEE and NATO radar frequency bands, with respective wavelengths [20]

Detecting Difficult-to-Detect Threats

The above computer simulation analysis proves what has been known for long: *stealth threats are not so stealth at lower frequencies*. In other words, the basic principles of RCS reduction, i.e., purpose shaping and special radar absorbent coating, are less efficient at lower frequency bands [19]. So, stealth airplanes or missiles are optimized for higher frequencies, from the S-band and above. Anyway, most dangers (i.e., fire control radars) are in these frequency bands.

Indeed, ground-based air defense system radars may emit in the S-band (for search) and in C or X-band (for tracking/fire control). Aircraft fire control radars operate in the X-band, while missile radar seekers may operate in the X or Ku-band. At lower frequency bands, there are mostly surveillance radars, which do not pose an imminent danger. A notable exception would be the Soviet-era radars used as search radars in anti-aircraft batteries, such as the P-18 which helped downing one F-117 during the war in Yugoslavia.

On the other hand, wide-band stealth is rather impossible or at least not cost effective. At lower frequencies, major aircraft parts, such as wings, horizontal and vertical tails, fall into the resonance or Mie scattering region, as their principal dimension becomes comparable with or a multiple of the radar wavelength. In this way, the primary effect of purpose shaping, which is to avoid scattering the incoming radiation back to the transmitting radar, is degraded. This is more evident on smaller planes, while large planes such as the B-2 stealth bomber are more "immune" to this phenomenon, namely they conserve their low observable characteristics even at low frequencies. This explains why the B-2 is a "flying wing": if it had any horizontal or vertical tails, it would enter more easily the resonance region, possibly increasing RCS, if illuminated by low frequency radiation [19]. Furthermore, RAM is inherently narrow-band and cannot be efficient at lower bands. In other words, much more thick coatings would be required in order to efficiently protect the skin at lower frequencies, increasing weight and cost prohibitively [18].

Low Frequency Band Radars

According to the above rationale, *low frequency band radars* seem to be a promising approach against stealth threats, including ballistic missiles. Such radars operate at frequencies in the L-band (1.2–1.4 GHz) or lower. Practically, apart from the L-band, low frequency radars operate in the UHF-band (~0.5 GHz) and VHF-band (~ 150 MHz). In this category, over-the-horizon radars operating in the HF-band should also be mentioned, based on either the surface wave principle or on tropospheric scattering.

Frequency and wavelength have an inverse relationship, so lower frequency means longer wavelength. This, in turn, would imply larger antenna, excessive volume and weight and limited transportability. Such a radar would be a perfect target on its own. Furthermore, low frequency radars are susceptible to clutter and cannot provide the necessary accuracy for fire control [21]. They are employed mostly for early warning, cueing higher frequency (and thus more accurate) radars to the direction of the target, increasing their probability of detection. Finally, the electromagnetic spectrum is quite congested at V/UHF, making difficult the allocation of unused frequencies for radar operation. For all these reasons, VHF/UHF radars have long been considered as obsolete in most Western countries and have been replaced by L and S-band radars.

However, having realized the significance of lower frequency bands, many countries (especially Eastern ones) have employed modern digital electronics technology to overcome some of the performance limitations inherent in V/UHF and other similar radar. With the progress of active electronically scanned array (AESA) antennae and improvements to computers and signal processing, lower-band radars have become more accurate and their range has increased [22]. Mobile low freq. radar systems have also been presented, which, despite their size, could be folded, getting ready to be transported in a few minutes.

Besides, low freq. radars cannot be detected or jammed by most aircraft self-protection systems, except for specialized wide-band ESM (Electronic Support Measures) and low band jammers, features not common to jet fighters. Especially V/UHF radars cannot be engaged by anti-radar missiles, such as the Raytheon AGM-88 High-speed Anti-Radiation Missile (HARM), or loitering munitions, such as the IAI Harpy. Therefore, low freq. radars offer some critical advantages, in addition to increased detection ranges against difficult-to-detect targets, including ballistic missiles, since quite a few of modern radars exhibit also ballistic missile defense capabilities.

A notable example of a complete radar system is the 55Zh6 M Nebo-M mobile multi-band radar complex, developed by the Russian Nizhny-Novgorod Research Institute of Radio Engineering (NNIIRT), which was the first to present a VHF AESA system. Nebo-M includes three truck-mounted AESA radar systems: the VHF RLM-M, the RLM-D in the L-band and the S/X-band RLM-S, as shown in Fig. 10.10. All three radars operate simultaneously, connected to a ground control vehicle, which performs data fusion. In this way, a target would be detected first by the VHF radar, which would cue the RLM-D. This in turn would cue the

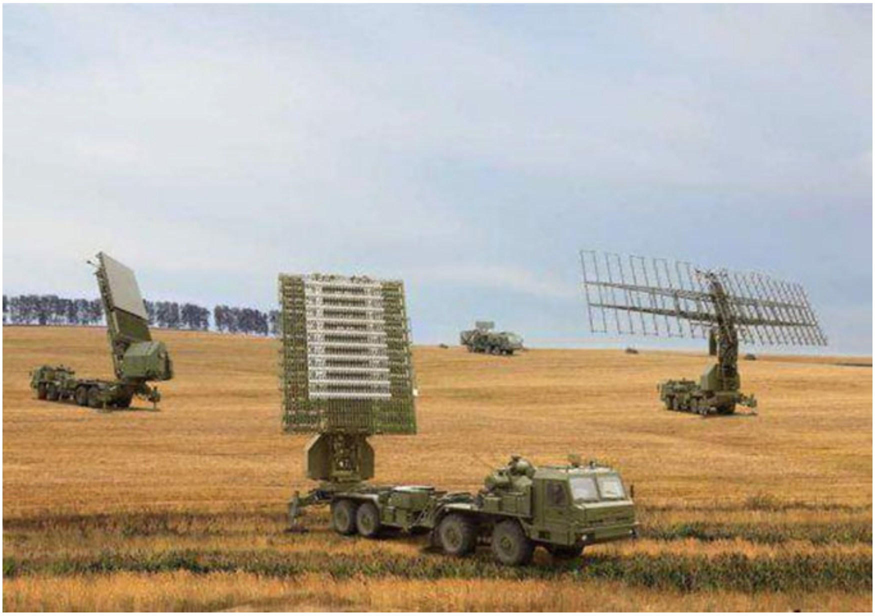


Fig. 10.10 The Nebo-M radar complex, comprising 3 radars in different bands (metric, deca-metric and centimetric, in terms of wavelength, or VHF, L and S/X, in terms of frequency) and a control station, performing sensor fusion [19, 23]

RLM-S, which could use a "stop and stare" technique, increasing the dwell time and thus the probability of detection, offering a weapon-quality track [22].

Recent examples of low freq. radars include the Thales SMART-L EWC (Early Warning Capability, featuring the latest GaN AESA technology), the IAI-ELTA UHF-band AESA ELM-2090U family, the CETC JY-27A Skywatch-V, as well as the HF IAI-ELTA ELM-2270 EZ GUARD coastal surveillance system, which is able to detect also low flying aircraft. Other examples can be found in [9, 21, 22].

Considering the example of the Alenia-Marconi L-band S743D Martello 3D surveillance radar, a detection range of 200 n.m. against a standard target of 1 m^2 can be assumed, taking into account a brochure claiming "long range detection of small, fast targets at distances beyond 200 nm". In the L-band, the F-35 RCS has been predicted to be in the -9 dBsm class, without RAM (see Fig. 10.5). Emulating the use of RAM and following a similar approach as previously, the attenuation in the L-band can be estimated to be 3 dB lower than the one in S-band (7 dB), that is 4 dB. In this way, the F-35 RCS is -13 dBsm or 0.05 m^2. Such a target would be detected by an S743D at approximately 95 n.m. Compared to the S-band HR-3000, the L-band S743D offers almost 50% more detection range against the F-35. Taking into account the RCS increase at lower frequencies and the narrow-band nature of RAM, equivalent V/UHF radars are expected to exhibit even longer detection range against the F-35, exceeding 100 n.m. Please note that this value is still small, compared to the detection range of typical targets (with RCS > 1 m^2), which exceeds 200 n.m.

Even if a range of 100 n.m. would be acceptable for early warning and fire control radar cueing, it should be noted that all above-mentioned range values pertain to electromagnetically clear conditions. If the environment is congested by strong electronic warfare transmissions (by specialized low band jammers), detection performance is degraded.

Taking into account the above, modern 3D AESA low frequency band radars with advanced digital processing and ECCM (Electronic Counter-Counter Measures) capabilities should be considered as the basic building block of an integrated air defense system, capable of countering difficult-to-detect targets, such as stealth aircraft, ballistic missiles, as well as cruise missiles and UAVs, depending on the radar coverage. Such radar network should be dense enough, keeping in mind a detection radius of less than 100 n.m. and sufficient radar overlap, as well as sensor redundancy. Mobility is also a key issue, for enhanced survivability: a fixed radar is a known target, with limited lifetime in time of war.

Passive Coherent Location (PCL) Radars

A network of low frequency radars as discussed above would sufficiently cover the given airspace against all kinds of threats, at least from a certain altitude and above, depending on the radar horizon. However, a very low-flying aircraft, cruise missile or UAV, exploiting coverage gaps due to ground obstacles (mountains, hills, islands, etc.) and remaining as long as possible in the radar shadow, could deceive the air defense system and approach dangerously close to an asset before being detected. One solution to this problem, albeit an expensive one, would be a very dense radar grid. Another idea would be the use of a different kind of sensor, to act as a gap filler.

In this context, a viable approach would be the use of *passive radars*. The operation of passive radars, also known as Passive Coherent Location (PCL) radars or Passive Bistatic radars, is based on the exploitation of existing transmissions. At all times, there are various transmissions (e.g., FM radio, DAB, analog/digital TV, HDTV, GSM, 3G), covering significant parts of the lower airspace. A passive radar comprises a "reference" antenna, directly receiving the broadcast of a station, and a "target" antenna, searching for a potential target. In case of a target being present, the signal from the station will be possibly received also by the "target" antenna, shifted in time (due to the longer distance covered to and from the target), shifted in frequency (due to the doppler effect, since the target is moving), and of course at a considerably lower power level (due to the longer distance and the scattering on the target). Therefore, if a signal similar to the direct signal of the "reference" antenna is received by the target antenna, there is a potential target, as shown in Fig. 10.11. Comparing the two signals and taking into account the relevant geometry (directions of antennae and relative position of the station), the position of the target can be calculated [24, 25].

Passive radars offer some certain advantages in the modern warfare, such as the following:

1. They provide covert detection and tracking.
2. They cannot be detected by aircraft self-protection systems or even more dedicated ESM (Electronic Support Measures) systems, they cannot be easily

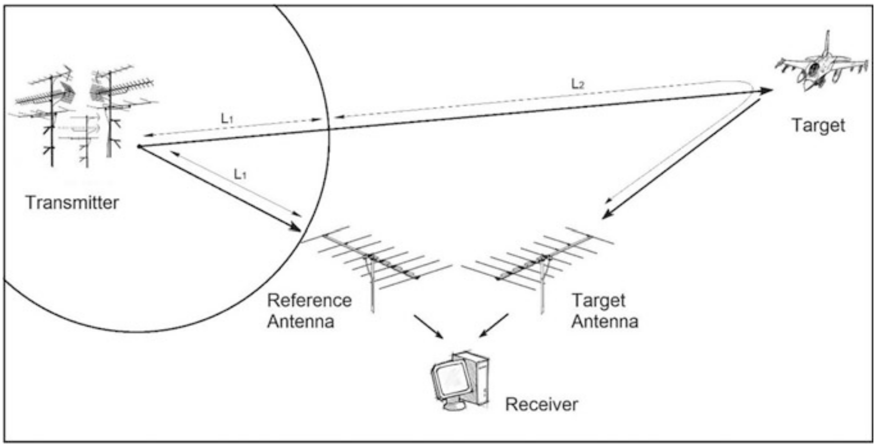


Fig. 10.11 The principle of operation of the passive radar: measuring the difference of the time of arrival of the "direct" signal from a station and the same signal after having being scattered on the target [25]

jammed, and they cannot be targeted by anti-radiation weapons (such as the AGM-88 HARM and the IAI Harpy).

3. They involve lower budgetary requirements, both for procurement and for operation (e.g., there is no transmission, so there is no need for expensive electron tubes and associated circuitry).
4. They typically involve transmissions in V/UHF, so they fall into the category of low frequency band radars, with the relevant anti-stealth capabilities, as explained in the previous Sections.
5. Furthermore, no license is required for their operation, as would be the case for active radars in congested environments, such as at the vicinity of an airport.

On the other hand, they present some drawbacks, such as the dependence on the geometry and on signals not optimized for radar use, the increased computational requirements, the inability to detect anything at higher altitudes (since there is practically no broadcast above 10000–15000 feet), and the difficulty to provide 3D tracking (many PCL radars are 2D).

Despite the various shortcomings, it seems that many countries are developing PCL radars, even if they may not admit to do so. There have been examples of PCL systems which were once announced, promoted for some time and subsequently disappeared. Notable examples are the Silent Sentry 2 by Lockheed Martin [26] and CELLDAR by BAE Systems and Roke Manor Research [27], in the US and UK, respectively. The only relatively mature passive radar is the Homeland Alerter 100 by the French Thales Air Systems [28]. More recently, a passive radar was proposed in Germany by Airbus Defence and Space (ex Cassidian), as well as the Italian AULOS Passive Covert Location Radar by Selex Sistemi Integratti. Now, the emergence of low-cost Software Defined Radios (SDR), as well as the abundance of cheap computers, have allowed the implementation of PCL systems, not only by radar manufacturers, but also by non-governmental agents, such as enthusiasts and students in electrical engineering.

The unique capabilities offered by the PCL approach in the context of the modern battlefield, especially against stealth and low flying threats, in combination with the covert operation and the low cost, make them a viable candidate for the gap filler role, in order to cover the lower tier of the airspace, depending of course on the availability of the existing transmissions.

The combination of a passive radar system with one or more dedicated transmitters, emitting a suitable, powerful signal, would transform the passive radar to a bistatic/multistatic system. In this way, issues like the coverage at higher altitudes or at areas without sufficient existing transmissions (e.g., over the open sea) could be mitigated. The transmitted signal could be disguised with a modulated content, like music. Even if the dedicated transmitters cease to emit (due to enemy attack or sabotage), the system would fall back on pure passive mode. At the present time, there are not so many multistatic radar systems available. However, such systems exhibit some serious advantages, including the backup passive mode, and should be investigated more thoroughly.

Finally, the category of ESM radars should also be mentioned, that is passive sensors/radiolocators which measure the time difference of arrival (TDOA) of pulses at three or four sites, in order to accurately detect and track aircraft exploiting their own emissions. Systems of this category require some kind of transmission from the target, such as IFF/SSR, TACAN/DME, radar or even jamming signals. If an intruder attacks silently, he cannot be detected by such a system. However, ESM radars offer an accurate and low cost means of surveillance, at least for everyday operations. This principle has been used in the frame of civil air traffic control, exploiting ADS-B transmissions, known with the term *multilateration*. In the defense context, if a potential intruder is aware of the existence of such sensors, he would have to apply even more stringent EMCON (emission control), imposing a further restriction to his activities. The most well known example is the family of the Czech Vera-NG. There is also a number of eastern systems, such as Kolchuga-M, VEGA 85V6-A and DWL002 [25].

Unbreaking the Kill-Chain

According to common practice, the "kill chain" from the point of view of the defender against an intruder comprises the following steps:

1. Detection (usually by a long range surveillance radar),
2. Identification (e.g., with the use of IFF/SSR),
3. Tracking (with a fire control radar, either ground-based or air-borne),
4. Weapon/Asset Selection (e.g., ground-based air defense or jet fighter),
5. Engagement (fire the selected weapon),
6. Assessment (evaluate the results of the engagement).

Detection is the first step towards any reaction against an intruder. That is why the discussion in the previous sections is focused mainly on the detection of enemy threats. Assuming now that a suitable radar network has detected a target and that this target has been identified as hostile (e.g., by failing to respond to IFF Mode 4/5 interrogation), the next step requires tracking by a radar which should be able to provide a weapon-quality track, operating typically in the C, X or Ku-band.

Using the F-35 as a case study, according to Fig. 10.5 and the relevant analysis, its RCS is very small in higher frequency bands, at the order of 0.01 m^2 or even less. Therefore, there is a high probability that even if an F-35 could be detected by lower frequency surveillance radars, it would not be detected by fire control radars, operating at higher frequencies. This is more probable in the case of aircraft fire control radars, where volume, weight and power limitations, impose restrictions on the radar power—aperture product. The situation is even worse concerning missile radars. More analytically:

a. According to open source info, the Northrop Grumman AN/APG-68(V)9 mechanically scanned array radar, equipping recent blocks of F-16 jets, can detect a standard target of 1 m^2 RCS approximately at 38 n.m. [13]. This radar would not perform well against the F-35, which exhibits an RCS of about 0.01 m^2 in the X-band, allowing an F-16 to detect it at as close as 12 n.m. In other words, an F-16 would get inside the air-to-air missile envelope of the F-35 before picking it up.

b. Concerning the F-35 radar, the AN/APG-81, by Northrop Grumman as well, open sources cite 150 km (that is 81 n.m.) against a 1 m^2 RCS target. Solving for 0.01 m^2, it seems that an F-35 can pick up another F-35 at a range of a little more than 25 n.m. Even if this distance is more than double the previously mentioned range of the F-16, still it is small, allowing the pilot to gain only a limited perception of the tactical situation against stealth threats.

A list of various radars and their estimated detection ranges against the F-35 can be found in [9] (with the F-35 RCS assumed to be equal to 0.0015 m^2, according to an unofficial USAF “leak”).

This issue has been known as “breaking of the kill chain” [22]: even if a stealth target, such as the F-35, is detected by surveillance radars, the track may not be possible to be handed over to jet fighter radars in order to intercept it. Moreover, even if a fighter achieves a missile launch against such a target, the missile may not “see” the target when it goes active. This applies to ground-based air defenses, as well.

In order to “unbreak” the kill chain, the following issues should be considered:

- AESA radars offer unique advantages compared to older, mechanically scanned array (MSA) radars. Perhaps the more obvious advantage is longer detection range (almost twice the range compared to MSA radars). However, it should be noted that as the beam reaches high off-boresight angles, the maximum detection range, as well as the accuracy, are degraded considerably. This issue does not affect MSA radars.
- Apart from the radar, almost all jet fighters now employ InfraRed Search & Track (IRST) systems. These are passive sensors which offer serious advantages, such as longer detection range and much better angular resolution with respect to the radar, while they cannot be jammed easily. On the other hand, they cannot measure distance accurately enough and their performance depends on the weather conditions. Newer generation IRST systems are advertised to exhibit anti-stealth capabilities [22].
- As shown above, the fighter aircraft radar is simply inadequate to provide effective situational awareness against stealth threats. Therefore, the “big picture” should be transferred from the Air Control System, via tactical data link (e.g., Link16) to the fighter aircraft. In this way, fighters would be aware of the all-around tactical situation, even without turning their radars on, preventing also the intruder from locating and identifying them on his RWR display.
- Furthermore, fighters should have the ability to engage a track transferred via data link, even if it is not confirmed by their own radar, possibly firing a missile in Lock-On After Launch (LOAL) mode. This implies lower Pk (kill

probability), since the track supplied via data link is of lower quality with respect to the aircraft radar track. However, this is better than waiting forever to get a radar track.

- The same applies to ground-based air defense systems, which they should also be networked, receiving tracks from the Air Control System and engaging them automatically, even without confirming the targets with their own radars.
- It is clear that no single sensor is able to cope with a stealth threat effectively. Different kind of sensors should be employed, covering multiple frequency bands, from HF and VHF to optical, and their readings should be correlated and fused.
- Netcentric warfare principles should be applied: information from every available sensor should be used, identified, fused and transferred to the shooter or even to the missile, in order to engage the target [29].

Conclusions

The air threat today is not limited to conventional, fast flying jet fighters and bombers. In the modern air battle, one can find stealth fighter aircraft, ballistic and cruise missiles, UAVs (Unmanned Air Vehicles) flying at various flight profiles, as well as heavily loaded fighters/bombers, with advanced avionics for discrete and accurate air-to-air and air-to-ground targeting. Furthermore, all these threats feature more or less reduced radar and infrared signatures. Legacy air defense radars may find it rather difficult to provide early warning against such threats, especially under strong electronic warfare transmissions.

In this context, using the Lockheed Martin F-35 stealth fighter as a case study, it was proven that low frequency band radars (i.e., L-band or lower bands) offer some significant advantages, especially radars operating in the V/UHF-band. Therefore, modern low frequency band AESA radars, with advanced digital processing, ECCM and ballistic missile defense capabilities can be used as a building block in an integrated air defense system. These radars offer detection ranges of the order of 100 nautical miles against targets, such as the F-35.

In order to fill gaps in the radar coverage, passive radars are proposed, which exploit existing transmissions (e.g., FM, TV, mobile telephony), providing coverage at low to medium altitudes. Furthermore, since they do not emit on their own, they cannot be detected and threatened by anti-radiation weapons.

Finally, some aspects of maintaining the “kill chain” were discussed, pointing out the importance of multi-band sensors, data fusion, tactical data links and network-centric warfare.

Survival against today’s air threats requires a suitable adaptation at all levels, a wide transformation of assets, an evolution of capabilities, in fact a transition to a new era. Failure to comply with these requirements would result to limited situational awareness and could lead to loss of assets without any prior warning.

References

1. Skolnik MI (2001) Introduction to radar system, 3rd edn. McGraw-Hill, New York
2. Skolnik MI (2008) Radar handbook, 3rd edn. McGraw-Hill, New York
3. Stimson GW (1998) Introduction to Airborne Radar (Aerospace & Radar Systems), 2nd edn. SciTech Publishing, North Carolina
4. Adamy D (ed) (2001) EW 101: A first course in electronic warfare, 1st edn. Artech House Radar Library, Massachusetts
5. De Martino A (2012) Introduction to modern EW systems (Radar). Artech House, Massachusetts
6. Du Plessis WP, Odendaal JW, Joubert J (2009) Extended analysis of retrodirective cross-eye jamming. IEEE Trans Ant Prop 57(9):2803–2806
7. Qu CW, Xiang YC (2010) Active cancellation stealth analysis based on RCS characteristic of target. Radar Sci Technol 8(2):109–112
8. Knott EF, Shaeffer JF, Tuley MT (2004) Radar cross section, 2d edn. SciTech Publishing, North Carolina
9. Zikidis K, Skondras A, Tokas C (2014) Low observable principles, stealth aircraft and anti-stealth technologies. J Comp Mod 4(1), 129–165. http://www.scienpress.com/download.asp?ID=1040. Accessed 21 July 2017
10. Limnaios G (2014) Current usage of unmanned aircraft systems (UAS) and future challenges: a mission oriented simulator for UAS as a tool for design and performance evaluation. J Comp Mod 4(1):167–188. https://www.scienpress.com/download.asp?ID=1041. Accessed 21 July 2017
11. Hassanalian M, Abdelkefi AB (2017) Classifications, applications, and design challenges of drones: a review. Prog Aerosp Sci 91:99–131. http://dx.doi.org/10.1016/j.paerosci.2017.04.003. Accessed 21 July 2017
12. Touzopoulos P, Boviatsis D, Zikidis KC (2017) Constructing a 3D model of a complex object from 2D images, for the purpose of estimating its radar cross section. J Comp Mod 7(1):15–28. http://www.scienpress.com/download.asp?ID=184991. Accessed 21 Jul 2017
13. Touzopoulos P, Boviatsis D, Zikidis KC 3D Modelling of potential targets for the purpose of radar cross section (RCS) prediction. In Proceedings of the 6th International Conference on Military Technologies (ICMT2017), Brno, Czech Republic, 31 May–2 June 2017, pp 636–642
14. Garrido E Jr (2000) Graphical user interface for a physical optics radar cross section prediction code. Master's Thesis, Naval Postgraduate School, Monterey, California. http://calhoun.nps.edu/bitstream/handle/10945/32958/00Sep_Garrido.pdf?sequence=1 Accessed 21 Jul 2017
15. Chatzigeorgiadis F Development of code for a physical optics radar cross section prediction and analysis application. Master's Thesis, Naval Postgraduate School, Monterey, California (2004). http://calhoun.nps.edu/bitstream/handle/10945/1453/04Sep_Chatzigeorgiadis.pdf?sequence=1 Accessed 21 July 2017
16. Jenn D (2000) MathWork– File Exchange — POFACETS4.1. https://www.mathworks.com/matlabcentral/fileexchange/35861-pofacets4–1. Accessed 21 July 2017
17. Shah TK, Malecki HC (2010) CNT-based signature control material. U.S. Patent Application Publication, US 2010/0271253 A1, Lockheed Martin Corp., 28 Oct 2010
18. Katz D (2016) The 'Magic' behind radar-absorbing materials for stealthy aircraft. Aviation week & space technology. http://aviationweek.com/aircraft-design/magic-behind-radar-absorbing-materials-stealthy-aircraft. Accessed 21 July 2017
19. Sweetman B (2013) Commentary: do russian radar developments challenge stealth? Aviation week network. http://aviationweek.com/defense/commentary-do-russian-radar-developments-challenge-stealth. Accessed 21 July 2017
20. Waves and frequency ranges. http://www.radartutorial.eu/07.waves/Waves20and20Frequency20Ranges.en.html. Accessed 21 July 2017

21. Katz D (2016) Physics and progress of low-frequency counterstealth technology. Aviation week & space technology. http://aviationweek.com/air-combat-safety/physics-and-progress-low-frequency-counterstealth-technology. Accessed 21 July 2017
22. Sweetman B (2015) New radars, IRST strengthen stealth-detection claims. Aviation week & space technology. http://www.aviationweek.com/technology/new-radars-irst-strengthen-stealth-detection-claims. Accessed 21 July 2017
23. Russia today: Russia deploying next-gen Nebo-M radar complexes to counter NATO threat https://www.rt.com/news/233959-russia-deploys-nebo-radars/ (2015) Accessed 21 July 2017
24. Griffiths H, Baker C (2013) Passive Bistatic Radar. In: Melvin WL, Scheer JA (eds) Principles of Modern Radar, vol 3. Scitech Publishing, North Carolina
25. Nomikos P, Economou D, Limnaios G, Zikidis K (2016) Presentation and feasibility study of passive radars. Air Force Rev Mag (in Greek) 107:86–103. https://drive.google.com/file/d/0B2Vf7Ad7I1njVFJiX3pTTlY1aU0/view . Accessed 21 July 2017
26. Baniak J, Baker G, Cunningham AM, Martin L (1999) Silent sentry passive surveillance. Lockheed Martin. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.475.8561&rep=rep1&type=pdf. Accessed 21 Jul 2017
27. Grossman WM (2003) Connect the pings, stealth radar from cell-phone radiation. Sci Am 26–28
28. Lok JJ (2007) Thales passive radar processes signals from radio towers. Aviation week. http://aviationweek.com/awin/thales-passive-radar-processes-signals-radio-towers. Accessed 21 July 2017
29. Zikidis K Low observable threats, RCS estimation and anti-stealth technologies. Paper presented at the 5th Air Power Conference, Hellenic Air Force, Dekelia Air Base, Greece, 16–17 Feb 2017. https://youtu.be/dFxb4U5b6no . Accessed 21 July 2017

Chapter 11
Mobile Data Fusion for Maritime Surveillance

Evangelos Sakkopoulos, Emmanouil Viennas, Zafeiria-Marina Ioannou, Vassiliki Gkantouna, Efrosini Sourla, Mersini Paschou, Athanasios Tsakalidis, Giannis Tzimas and Spyros Sioutas

Abstract Maritime surveillance operations are needed worldwide to monitor and reassure safety and security across the seas. Numerous devices are employed in order to provide situational awareness of the vast sea. Lots of different technologies are involved to provide multiple views and clarify maritime conditions at a given time and place, however making interoperability a real challenge. The task is even more tedious as there is a key request to provide a single window view for multiple even all possible inputs. In this work we present an integrated mobile fusion solution for multiple tracking and monitoring sensors (e.g. low weight/high performance radar, position transmission mechanisms and electro-optic/systems and hyper-spectral sensors) to assist the detection and early identification and tracking of moving targets (e.g. with moving target indication and data fusion/correlation capabilities), as well as methods for obstacle detection and maritime surveillance. This innovative single window mobile platform presents high efficiency, low operational costs profiles and contributes to standardization in construction as it utilizes typical tracking infrastructure and standard smartphones and tablets.

Keywords Mobile applications · Surveillance · Safety · Security

E. Sakkopoulos (✉) · E. Viennas · Z.-M. Ioannou · V. Gkantouna · E. Sourla ·
M. Paschou · A. Tsakalidis
Computer Engineering & Informatics Department, School of Engineering,
University of Patras, Rio Campus, 26500 Patras, Greece
e-mail: sakkopul@ceid.upatras.gr

G. Tzimas
Computer & Informatics Engineering Department, Technological Educational
Institute of Western Greece, 26334 Patras, Greece

S. Sioutas
Department Informatics, Ionian University, 49100 Corfu, Greece

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_11

Introduction

Orchestrating maritime surveillance operations and safety collaboration has a number of obstacles to overcome in terms of coordination using ICT and surveillance technology. Operations at sea include additional limitations than in typical continental cases. A number of typical technologies are employed to detect maritime incidents, the location, the direction, the historical positions and overall status of vessel traffic. The Automatic Identification System (AIS) is an automatic tracking system used on ships and by vessel traffic services (VTS) for identifying and locating vessels. It allows data exchange between ships and nearby AIS base stations. Data transmission is also possible through satellite AIS. AIS information supplements standard marine Radar, which continues to be one of the most important methods of collision avoidance and position detection for water transport. AIS main use is for traffic management, search-and-rescue, and banned vessel tracking. AIS information provides position coordinates of a vessel and therefore it is possible to overlay these positions on a Radar screen view in order to verify vessels detected and vessels without AIS transmission (e.g. smaller boats or banned vessels). This only shows the key importance and need for a single window data fusion system that may allow the overlaying ad hoc and systematically multiple sensors input.

The European Maritime Safety Agency (EMSA) operates and early warning system for significant maritime safety, rescue and security incidents in European water and other parts of the world. The key aims are to allow policy decision making on the European level and to accelerate the mobilization of European Union authorities to respond to oil pollution situations. The system also generates information to face other types of incidents.

However, AIS based tracking is far from a universal tracking solution. AIS mechanisms cannot securely—strongly identify a ship. AIS transmitters is possible to be switched off, or blocked from transmitting at vessels intending to remain unidentified and track-less that are considered a maritime security breach. Additionally, it seems possible to maliciously manipulate the AIS transmission by altering the GPS location and transmit mock positions of the ship and its route proving hard to detect with a number of possibilities for inappropriate use. Furthermore, small boats do not have the obligation at all to carry AIS mechanism. Therefore there are numerous ways that permit for unidentified vessels to move people and to transfer goods between neighboring countries and regions completely untracked. As a result, in such cases, discovery and rescue operations and identification procedures need to be conducted by sea vessels having visual contact and alternative tracking means such as radars, which however are not integrated into a single window system by design.

Thermal and optical cameras are also used to assist towards identifying position of ships and status of vessel traffic, especially at an incident location. GPS positioning is also critical for reporting the location of land and marine objects and staff.

The last but not the least of tracking devices are radars that may be found established at certain key surveillance points and also portable on vehicles and ships.

Combining all the aforementioned positioning technologies over a single information system screen has been an open challenge. Data fusion of multiple data streams is a tedious task. Additional obstacles include lack of adequate telecommunication infrastructure as there is only partial coverage at the open sea and even in several locations around a harbor area. Our approach provides a viable and solid solution for such data services and furthermore delivers for the first time to the author's knowledge the outcome of the single screen view of all crucial data combined to the vessels and vehicles operators. Traditionally, data are delivered as a whole but not fused together (with the exception of AIS over Radar) in operation centers of governmental agencies such as the coastguard and the navy. The operators receive commands and information over voice transmissions only.

A different longstanding problem in the shipping industry is the complexity and time involved in submitting reports when arriving in and departing from ports. This situation is related though to data fusion and single window need described above. In particular, ship operators, owners and agents are still burdened with having to fill in paper documents which include similar information and to distribute them to different government authorities, including port, maritime, safety, security, customs, border control, and health authorities [1]. A reporting gateway for the shipping industry: The Common Reporting Gateway module offers a comprehensive and harmonized interface (PC, tablet, XML) for the fulfillment of reporting formalities. This system may be linked with existing national back-end systems. Certain parts will also be made available to other Member States via SafeSeaNet [2]. In order to simplify this administrative processes the National Single Window (NSW) application has been designed. The NSW system assists for all maritime information to be reported in a single step by the ship data providers, at either national or port level, and made available to all relevant authorities. However, this national single window application provides solely administrative information and serves merely as a unified submission form.

Our approach is far more advanced as we have designed and deployed a complete monitoring single window solution for maritime surveillance. The proposed solution sends and receives multi modal mobile data at all participants. The data are not just raw data or combined raw data from Radar as in most cases, but they are being dynamically fused into a single multi layered window. In this way, we provide a single and combined information source for the safety of the maritime officer, the officials on the ground and on the ships authorized to access it.

An important advantage of the solution introduced uses the wide penetration of typical smartphones in order to provide to all parties involved with additional tools that will make their work easier and more efficient. As network interconnection cannot be taken for granted and while maritime operational staff is at sea, there has to be available a number of different ways to communicate with the parties involved in vessel traffic services. The better the communication and the exchange of data the easier the work of those involved becomes.

Smartphone applications may use limited resources, compared to desktop systems; however, by seizing a number of advantages such as portability, Internet access, location detection services, etc. mobile phone applications can provide advantages that ensure the results of the process. In our work, we present an analysis for efficient Mobile User Interface Design approach for maritime surveillance using Geo-informatics to combine multiple monitoring inputs and provide easier and more efficient decision support on the spot of the operations with typical smartphones and tablets.

Information provided by AIS equipment, such as unique identification, position, course, and speed, can be displayed on a screen or an Electronic Chart Display & Information System (ECDIS). AIS is intended to assist and allow maritime authorities to track and monitor vessel movements [3]. In cases of irregular or unusual circumstances there has to be a number of different ways for the parties involved in vessel traffic services to communicate. The better the communication and the exchange of data is the easier the work of those involved becomes. Smartphone applications may use limited resources, compared to desktop systems; however, by seizing a number of advantages such as portability, Internet access, location detection services, etc. mobile phone applications can provide advantages that ensure the results of the process.

The development of mobile services has already shown that it is possible to provide a simplified and harmonized mobile app for the maritime sector, enabling a smooth and fast flow of multi modal positioning, safety related operational data—initial shorter presentation of such solution at Sakkopoulos et al. [4].

The identified benefits include: For the shipping industry and the safety and security officials (a) User friendly and facilitates data inputs (b) Versatile: may be accessed via smartphone, tablet and PC with a common look and feel interface (c) Easier and quicker distribution of notifications to authorities (d) Re-use of information submitted in previous operations and historical analysis and data fusion (e) Easy-to-view decisions for authorities, (f) ability to communicate with multiple data providers onboard ship or vehicle, (g) Consolidation of information provided by different data providers for a single port call, region or country view.

The paper is organized as follows: section "Related Work" discusses related work. Section "Software Prototyping" discusses the software prototyping techniques that were applied and the advantages stemming from the procedure. Section "Integrated System: Architecture" presents the architecture and design of the integrated system. Section Mobile Application for the Captain and Crew describes the mobile applications of the integrated solution in detail. The last Sect. Mobile Application for the Members of the OperationalTeam discusses the results and concludes the paper.

Related Work

Vessel Traffic Monitoring

In EU waters SafeSeaNet [5] has been set up since 2010, which is a vessel traffic monitoring and information system, established in order to enhance maritime safety, port and maritime security, marine environment protection and efficiency of maritime traffic and maritime transport. It covers all European coastal waters (over 20,000 vessels), with 12,000 ships/day tracked in EU waters and 100,000,000 AIS positions recorded per month.

It has been set up as a network for maritime data exchange, linking together maritime authorities from across Europe. It enables European Union Member States, Norway, and Iceland, to provide and receive information on ships, ship movements, and hazardous cargoes. Some of the main information elements that are contained in the system and made available to users are the following ones:

- Automatic Identification System (AIS) based near-real-time ship positions (i.e. one every 6 min)
- Archived historical ship positions (over several years)
- Additional information from AIS-based ship reports (e.g. identification name/ numbers, flag, dimensions, course, speed, dimensions, destination and ship type)
- Estimated/actual times of arrival/departure
- Digital map layers (containing information on depths, navigation aids, traffic separation schemes, anchorages, AIS station locations, etc.).

However, vessel traffic monitoring system does not combine and integrate additional mobile information sources other than AIS. Our proposed solution provides bi-directional communication with the mobile service holders of our system to deliver and post positioning information from multiple sources beside AIS and map layers (e.g. radars, cameras, thermal cameras, GPS etc.).

EU Electronic Solution for Simplifying Administrative Procedures

Directive 2010/65/EU establishes that EU Member States shall accept the fulfillment of reporting formalities in electronic format and their transmission via a single window no later than 1 June 2015. Similar approaches have been issued for the world-wide waters. This single window shall be the place where all information is reported once and made available to various competent authorities and other Member States. To fulfill the directive, a recent prototype (2/2015) supports the form submission of the following formalities referred to in Directive 2010/65/EU:

- Notification for ships arriving in and departing from ports of the Member States (Directive 2002/59/EC),
- Border checks on persons (Regulation (EC) No 562/2006),
- Notification of dangerous or polluting goods carried on board (Directive 2002/59/EC),
- Notification of waste and residues (Directive 2000/59/EC),
- Notification of security information (Regulation (EC) No 725/2004),
- Entry Summary Declaration (Regulation (EEC) 2913/92 and Regulation (EC) 450/2008).

The National Single Window (NSW) [1] aims to simplify the administrative burden by providing a place where all maritime information is reported once by ship data providers, at either national or port level, and made available to all relevant authorities.

However, our proposed solution is a set of mobile data fusion services that are published at a single window UI across several different devices for PC, smartphone and iPad/tablet using the very same look and feel. The key feature of the services is that they provide access to fused data sources in a bi-directional manner and allow sending and receiving positioning and safety information from multiple mobile sources (e.g. radars, cameras, thermal cameras, GPS, AIS etc.).

Software Prototyping

Software Prototyping [6] is the process of creating prototypes of a software application, i.e. incomplete versions of the software program that is being developed. For the proposed system this procedure was particularly important because the mobile applications were designed for use under unusual and sometimes difficult circumstances faced in marine operations. Software prototyping has been applied to the process of developing the new software, taking into consideration the feedback from potential users. The process of software prototyping has many advantages and benefits, such as:

- Designers and software developers can receive valuable information from users at the beginning of the project.
- The client and the contractor can compare if the software matches the specifications of the software, which were followed for the development of the software.
- It allows the software engineer to have some insight into the accuracy of the initial estimates of the project and whether the deadlines and milestones proposed can be met successfully.

The primary purpose of the prototype is to enable users to evaluate the proposals of developers for the design of the end product with potential real test, rather than the interpretation and evaluation of the design based on descriptions only. Prototypes can also be used by end users to describe and demonstrate the requirements that have not been included, which may be a key factor in the trade relationship between the developers and their clients. This is the main purpose of using interactive design.

Key Features for Creating Prototypes

The key features for creating successful prototypes include supporting creativity by helping developers to have new ideas, to facilitate the exploration of design and to disclose relevant information regarding the users and their work [7]. Additionally, creation of prototypes should facilitate communication, by helping designers, engineers, managers, software developers, customers and users to discuss the options and to interact. Timely evaluation should be allowed, since they can be tested in many ways, including traditional usability studies and informal feedback, throughout the design process.

For the particular system presented here, the software prototyping procedure included the features presented above and resulted in the adjustment of the user interface to meet the particularities and unique requirements of people who work in ships or in the shipping industry in general. These end-users may have to face unexpected conditions and limitations posed while being at sea, thus they request a user interface friendly and easy to use, having all frequently used buttons and controls positioned within easy access (e.g. for a mobile or tablet application these controls have to be placed to the right or left side of the screen). The results of the software prototyping procedure were taken into consideration for the design and deployment of the individual parts of the proposed system.

Strategies of Prototypes' Development

The main four different prototypes' development strategies are [8]: (1) horizontal strategy, (2) vertical strategy, (3) strategy oriented to tasks and (4) strategy based on a scenario. Each strategy focuses on different design issues. For the purposes of this work Justinmind Prototyper [9] was used, which is a tool for creating software prototypes and high-fidelity wireframes of web pages. Concluding the prototype design and development, as a next step, a cross-platform mobile and desktop approach has been used to deliver the solution to the widest range of devices possible.

Integrated System: Architecture

This section presents the architecture of the proposed system, both in functional and physical level, which resulted using the functional requirements of the system and the software prototyping techniques presented in the previous section. Additionally, we determine here the interactions among the various subsystems' services of our integrated solution.

The system architecture covers early warning data fusion services, safety and surveillance services and information sources. The entire solution is based on a common look and feel user interface and provides assistance and information to its users in a streamlined manner. Early warning section, includes services and data sources:

- AIS historical data and analyzed information of ship tracked position.
- Digital map layers of historical incidents and maritime discovery and rescue operations in order to detect similarities or deviations with the operation at hand.
- Data fusion services that produce and communicate alerts and warnings for incidents to all users either in the operational centers or mobile on vehicles and ships.
- Textual and multimedia historical information that serves as logging for all previous operations to ensure personnel safety and re-assure high operation standards.
- Earth-Observation historical data.

The mobile services that are real time are coupled to the desktop web services through common communication architecture. The near/real time architecture of our system includes the following key services and subsystems:

- Digital Integration for multiple monitoring and surveillance sources such as thermal, optical and infrared cameras, radars, and real time AIS positioning, voice and data GSM/CDMA and SAT communications, Earth Observation incoming satellite data. Similar needs for integration and data transmission are met in other domains, e.g. e-health [5].
- Real time communication with all desktop users through multiple channels (e.g. land lines, SAT, GSM/CDMA, wifi direct, SMS text messages-asynchronous mobile service). An approach for real-time communication is presented in [10].
- Near real time communication with all mobile users that operate on terrestrial, coastal or maritime locations.
- Single window—common UI interface for all with roaming profiles that ensure that each user has all his/her data whatever the devices that has in front of him/her at the office, at the operation center, or onboard a vehicle or a ship.
- Offline service provisioning to allow making the most out of the proposed system while being at rural, coastal and maritime areas with extremely low telecommunication coverage (e.g. partial GSM/CDMA, low Sat communications, RF channels).

- Mobile Delivery Service to setup preinstalled offline data on mobiles. For the mobile users, all data are stored locally on encrypted mobile database to ensure availability of historical fused data for the operation area, anytime [11]. Data includes offline maps (OpenStreetMaps, nautical maps, multiple maritime information layers, historical operational layers, fused tracking information data etc.). Google maps may also be served for offline (with limitations of 50 × 50 km map area and 30 days maximum offline use, according to Google terms of use).
- Mobiles Synchronization Services that exchange information from and to mobile users through any available channel in an asynchronous manner.
- Encryption and strong authentication service for all services and users.
- Voice and multimedia communication logging services.
- Mobile positioning service that allows the exchange of location data coming from GPS.
- Mobile service for data fusion that allows transferring location information derived from vehicle radars, optical, thermal and infrared cameras just as lightweight textual geo-location, minimizing data bandwidth demands, which constitutes a big challenge in mobility analytics [12].

The proposed integrated solution is composed of three main roles which work together to meet the needs for all actors involved at operations. The key roles supported are: (Fig. 11.1)

- Management and Operational mobile role for the captain and permanent crew of operational vessel

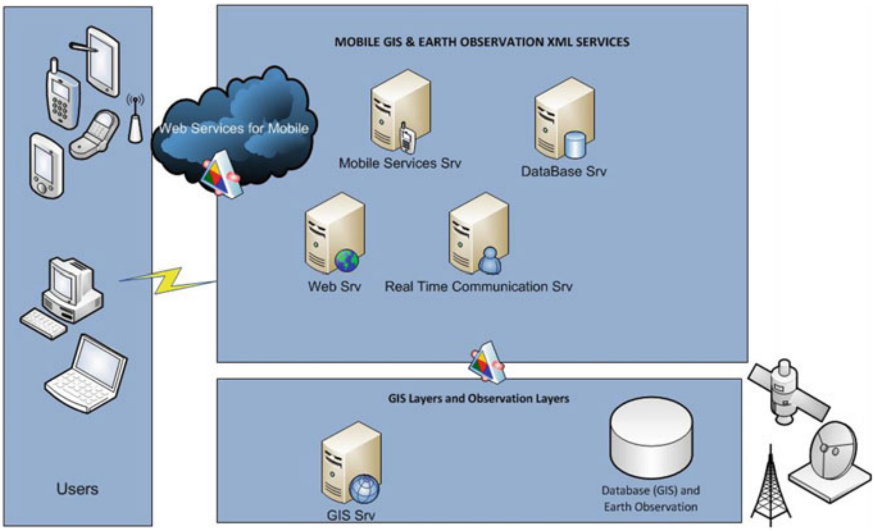

MOBILE GIS & EARTH OBSERVATION XML SERVICES
Web Services for Mobile
Mobile Services Srv
DataBase Srv
Web Srv
Real Time Communication Srv
GIS Layers and Observation Layers
GIS Srv
Database (GIS) and Earth Observation
Users

Fig. 11.1 Architecture of the proposed integrated mobile system

- Operational mobile role for the members of the operational team
- Management and Information role for the general staff of operational center and the respective authorities such as the Ministry of Maritime Affairs and EU headquarters' liaisons.

These roles are supported by respective PC and smartphone and tablet/iPad applications which are presented in detail in the following paragraphs.

The technological framework for the development of the solution includes:

- Open source mobile services running on the devices to facilitate functionality, communications and data exchange and provide data fusion and positioning information from mobile sensors (e.g. mobile radars, cameras onboard-on vehicle, GPS, etc.).
- Open source backend XML Web Services and data fusion services for digitization of stable sensors (e.g. radars, harbor cameras, AIS, etc.).
- Cross platform delivery approach using the Xamarin to create native iOS, Android, Mac and Windows apps. Xamarin apps look and feel native because they are. Xamarin apps leverage platform-specific hardware acceleration, and are compiled for native performance. This can't be achieved with solutions that interpret code at runtime.
- Open source GIS platform using Geoserver and Mapserver Services.
- Multiple custom services to allow data fusion on central historical database systems.

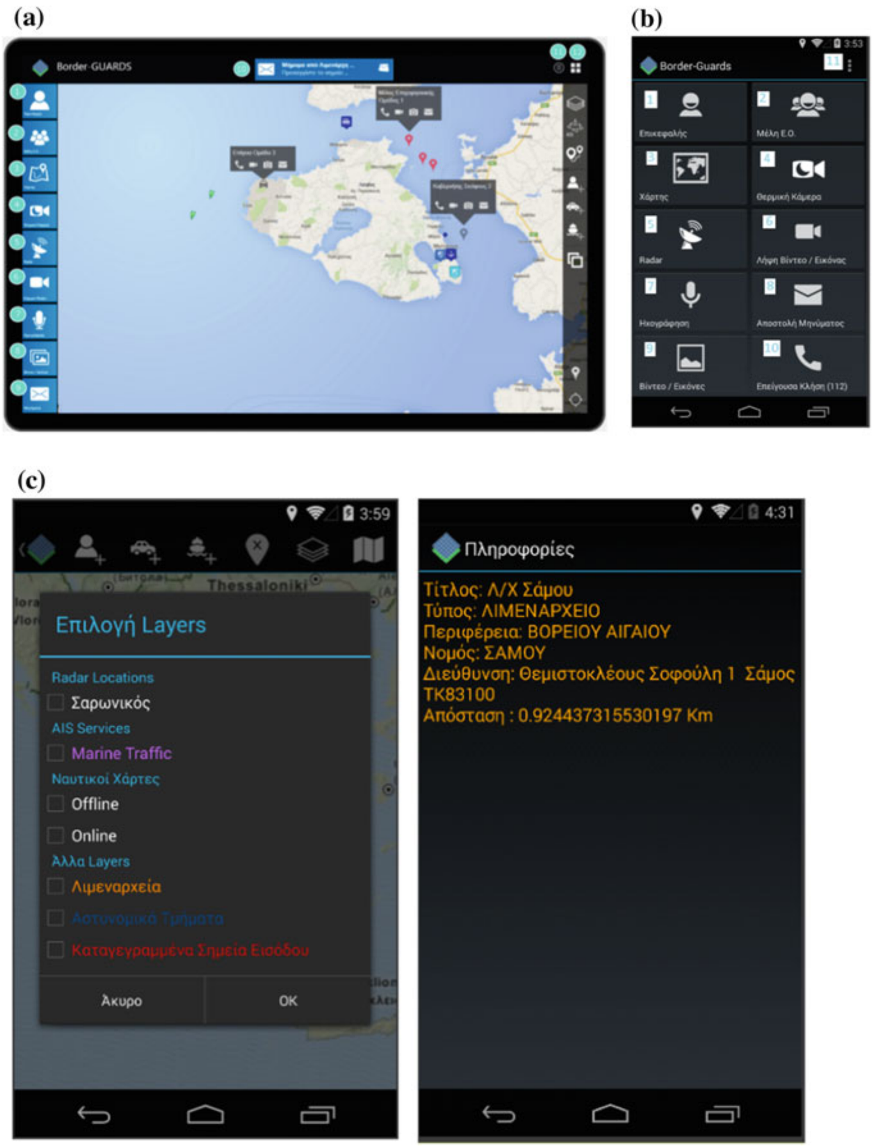
Mobile Application for the Captain and Crew

This application relates to the captain and the permanent crew located in the operational vessel. The main functions of this application are:

- Contact the harbormaster that coordinates the current operation.
- Contact the members of the operation through acoustic call, video call, sending images/videos or messages.
- View map and AIS information.
- View the location of the members of the operation.
- View image taken by a thermal camera.
- View image from radar.
- View image from a camera mounted onboard.
- Voice recording capability.
- View image gallery/video.
- Send/receive messages.

A. Functionality on the main screen (Fig. 11.2) includes:

- Harbor Master: Ability to communicate with the Harbor Master.



Τίτλος: Λ/Χ Σάμου
Τύπος: ΛΙΜΕΝΑΡΧΕΙΟ
Περιφέρεια: ΒΟΡΕΙΟΥ ΑΙΓΑΙΟΥ
Νομός: ΣΑΜΟΥ
Διεύθυνση: Θεμιστοκλέους Σοφούλη 1 Σάμος
ΤΚ83100
Απόσταση : 0.924437315530197 Km

Fig. 11.2 Main screen of the application for the captain and permanent crew in **a** tablet and **b** smartphone view **c Smartphone view**: layer selection, map selection, offline and online nautical base maps and layers. Colored layers indicate classes of points of interest in the region on ground or sea. POI details are shown in the same color to provide direct correlation and ease of use (on the right). **d Smartphone view**: multiple view selection possible (listview and large icon view (selected). It is also available language option selections for multi-national cooperation. On the right graphical map base is showing the Lesvos Island on the Aegean Sea in Greece, EU. **e Smartphone view**: offline detailed map base is showing details in the Lesvos Island on the Aegean Sea in Greece, EU. On the right hand side, a radar input is layered over the base map to show detected ships and airplanes in the region of Athens, Attica, Greece, EU

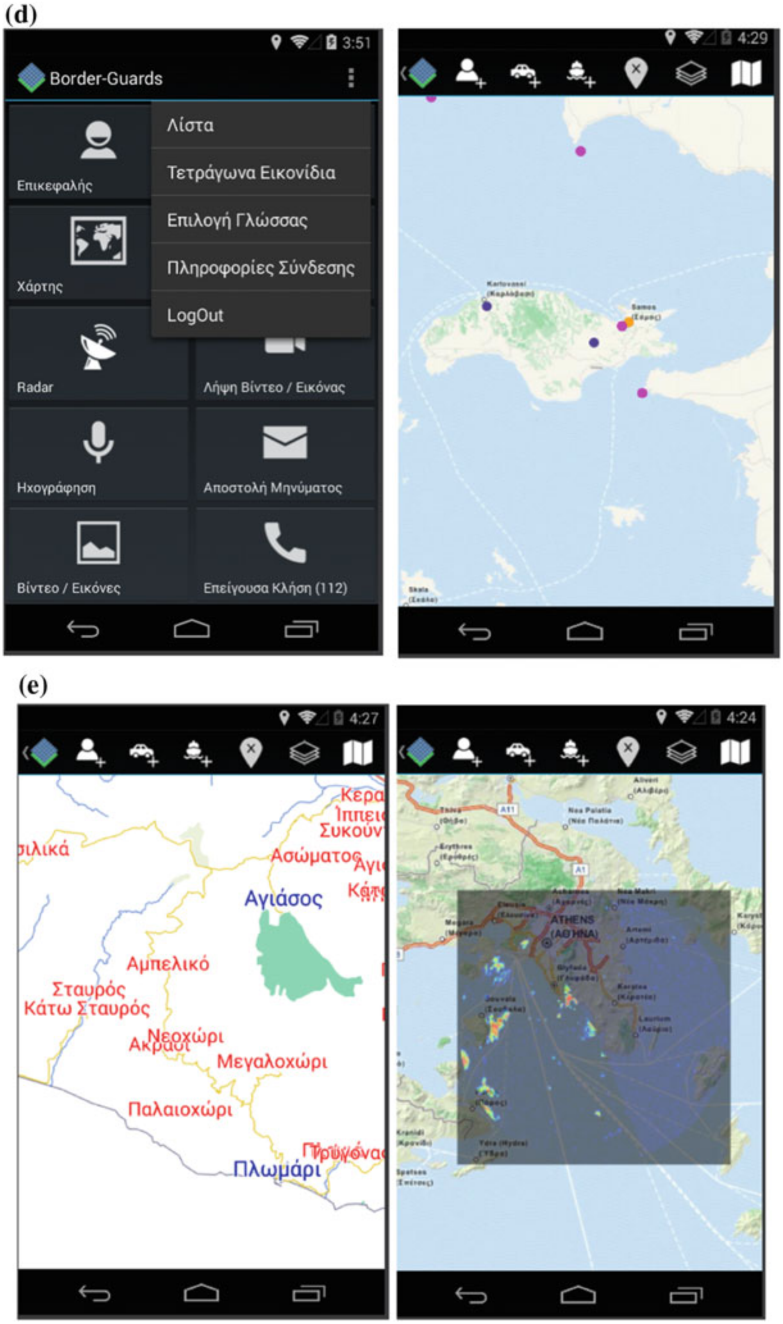


Fig. 11.2 (continued)

- Members of the operation: Ability to communicate with the members of the operation. By selecting this functionality the user is redirected to the “Members of Operation” part of the application.
- Map—AIS: View of the map—AIS data on the main screen.
- Thermal Camera: View image from thermal cameras, on the main screen.
- Radar: View image from radar, on the main screen.
- Camera Ship: View image from camera mounted onboard, on the main screen.
- Record: Ability to record conversations of the captain and the permanent crew of the vessel.
- Video/Pictures: View image gallery and video. This option redirects the application to display the “Collection Video/Image” screen.
- Messages: capability to send messages to the Harbor Master and the members of the operation.
- Notifications: Show notifications for incoming messages or pictures/videos.
- Message Window: Show a window with the latest incoming messages.
- Grid View: View “grid view” map, thermal camera, radar and view from the onboard camera to the main application screen.

B. Grid View Feature Functionality

As far as the “Grid View” option is concerned, the user may simultaneously observe feedback from: (a) the map (Fig. 11.2—point 1), (b) the thermal camera (Fig. 11.2—point 2), (c) the radar (Fig. 11.2—point 3), (d) the camera board (Fig. 11.2—point 4). Using option 5 of Fig. 11.2 (Map Icon) the user can reset the map in full screen mode of the application. With options 6, 7, 8 and 9 (Fig. 11.2) we can achieve full screen map display, thermal camera, radar and camera board respectively in the main application screen. As far as the functionality is concerned, the application will display the map of the surveillance area of the vessel captain. In addition, the following capabilities will be provided:

- Layers: Selection of map type and the displayed information (wind, harbors, marinas, etc.).
- AIS: Select to view information about the movements and ships’ positions based on data AIS, navigation data and data of ships and ports.
- Route: View the recorded route of the ship.
- New point (human): Insert a new human-point and relevant information.
- New Point (vehicle): Insert a new vehicle-point and relevant information.
- New point (ship): Insert a new point - the ship and relevant information.
- New window: Open a new window with a map.
- Groups’ Positions: Show/Hide the status/position of groups.

The map (Fig. 11.2a) provides the ability to view the position of the members of the operational team that the captain of the vessel oversees (icons in red) as well as members of another operational team (icons in green).

Choosing the position icon, the commander can see which member of the operational team is in the specific position and to select one of the following methods of communication:

- Acoustic call
- Video Call
- Uploading a photo
- Sending Message.

Additional Views are also supported to provide the best view for multiple smartphone screens and dimension compatibility. There is a list view possible as well as a large icon view for all available functionality. Furthermore, the solution includes full multilingual support to make it possible for use under multi-national cooperations in search and rescue incidents, exercises or security operations governed for example by alliances such as NATO or EU CoastGuard (former FRONTEX).

C. Data AIS—Functionality

The map of the application will display information about the movement and position of the ships, based on AIS data, navigation data, and data related to the vessels and ports. The settings for the displayed AIS information are managed using "AIS" (point 1), as shown in Fig. 11.3 and include the following options (point 2):

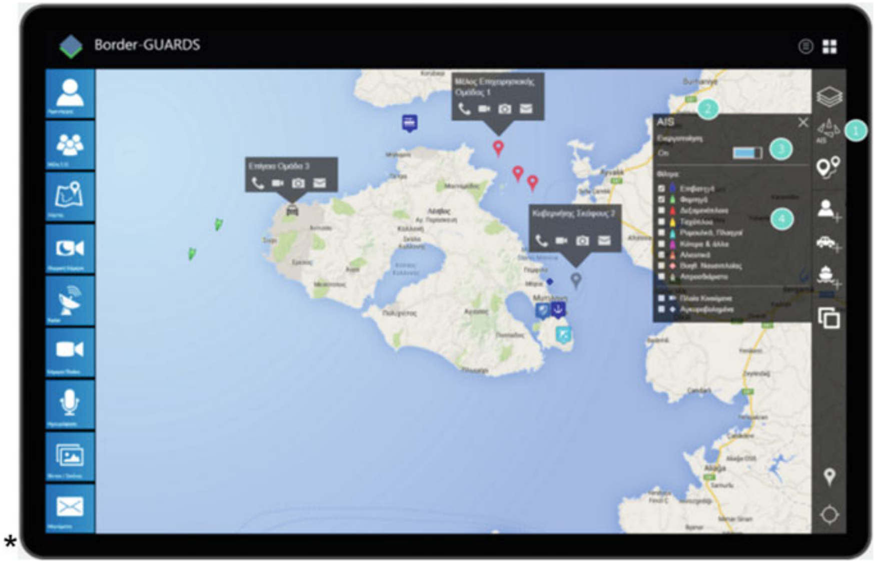


Fig. 11.3 Map and data mobile service at Lesvos island, Greece

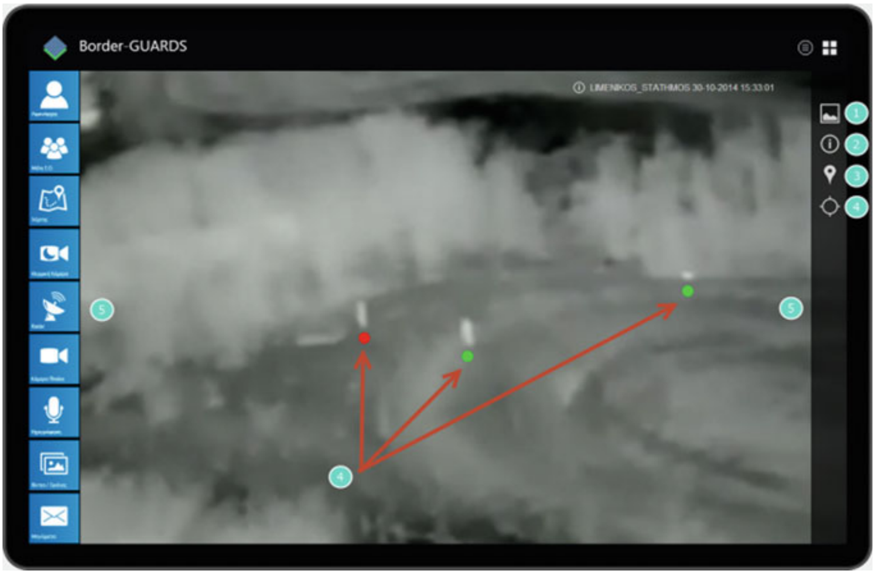


Fig. 11.4 Thermal camera augmented with friend (green) and foo (red) geolocation points

- AIS Enable/Disable (point 3).
- Filters for displayed information for different types of ships (point 4) e.g. Passenger ships, cargo, moving or anchored vessels.

D. Thermal Camera Functionality

The options provided to the user by the Thermal camera feature of this mobile application (as presented on Fig. 11.4) are:

- Saving and sending the image of the thermal camera
- Showing/Hiding information (e.g. station name, date and time)
- Insert Point of Interest (POI) in the image of the thermal camera
- Displaying the locality of points that represent the position of humans.
- Alternation of images (right–left).

E. Radar Functionality

The main functions of the radar functionality include:

- Saving and sending images from the radar
- Showing/Hiding information (e.g. station name, date and time)
- Introduction of points of interest in the radar image
- Alternation of images (right–left).

The options provided to the user of the radar feature of this mobile application (as they are presented and numbered on Fig. 11.5) are:

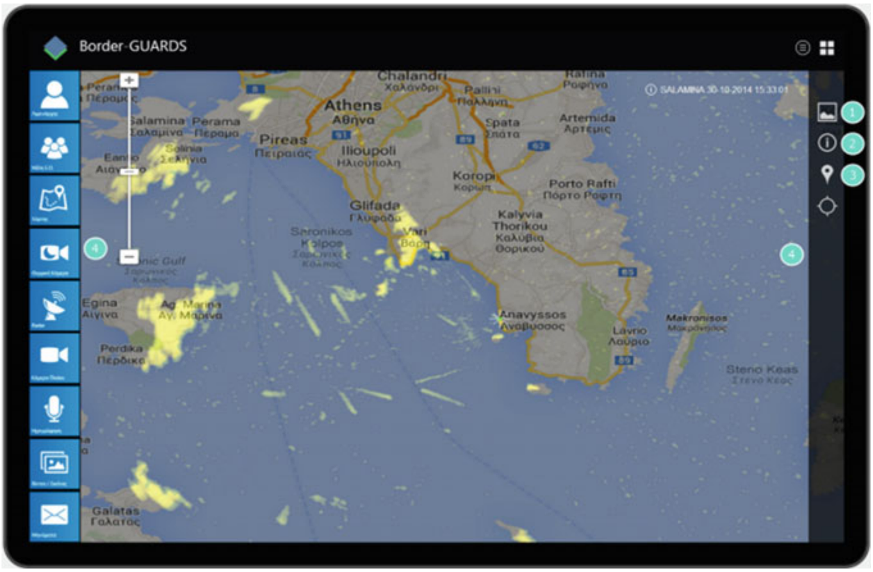


Fig. 11.5 Real time radar mobile service for region Athens, Greece

- Listing the members of the operational team
- Scope Search member of the Task Force
- Selecting a member from the list, the user can view information about the member and choose one of the following methods of communication:

(a) *Acoustic call*
(b) *Video Call*
(c) *Expanding the map for member positioning*
(d) *Sending an image*
(e) *Sending a message*

- By selecting “All Members” the captain can communicate quickly with all members of the operational team (e.g. by sending a message or an image).

Mobile Application for the Members of the Operational Team

This smartphone application is designed for use by the members of the operational team, who will be equipped with mobile devices. The main functions of this application include:

- Contact with the captain of the operational vessel.
- Contact with other members of the operational team through acoustic call, video call, sending images/videos or messages.
- View of the map and AIS information.
- View of the position of the members of the operational team.
- View image of the thermal camera.
- View image from radar.
- Image receiving and video recording capabilities.
- Voice recording capability.
- View of image/video gallery.
- Sending quick messages from a prepared list.
- Receiving messages with ability for automatic reading—recitation.

F. Main Screen Functionality

The options for the user of this mobile application (as presented on its main screen—Fig. 11.6a) are:

- Captain: Ability of communication with the captain of vessel.
- Members of the Operational Team: Ability to communicate with other members of the operational team.
- Map: View of the map to locate the position of the members and of other points of interest.
- Thermal Camera: View images from the Thermal Camera.
- Radar: View images from the Radar.
- Video/Image Downloading: Image capturing and video recording capabilities.
- Sound Recording: Voice or chat recording capability.
- Message Sending: Sending a message to members of the operational team or the captain.

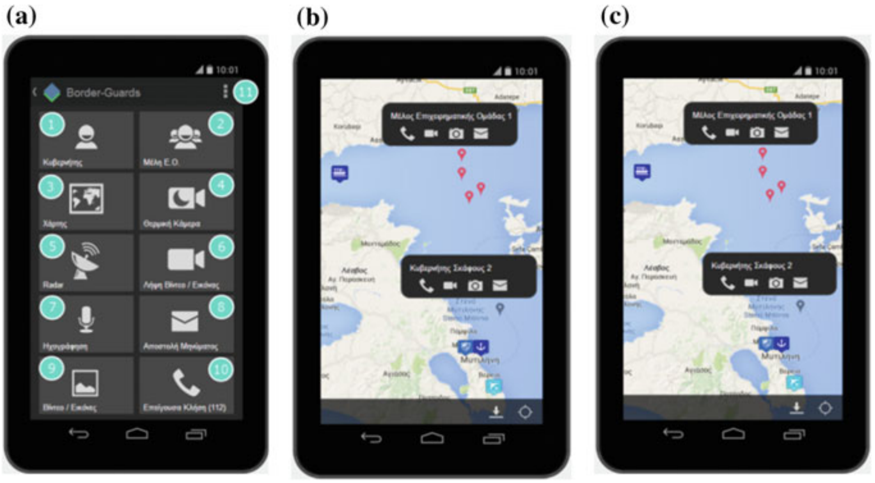


Fig. 11.6 **a** Main screen of the application, **b** map, **c** map downloaded for offline use

- Video/Pictures: Gallery with pictures and videos.
- Urgent Call: making an automatic call to the emergency number 112.
- Settings: Viewing and Language Settings.

G. Map Feature Functionality

The *map feature* (Fig. 11.6b) provides the user with the ability to view the position on the map of the members of his/her operational team (red spots on the map) as well as the members that belong to another operational team (gray spots on the map). By selecting a position spot, the user can view which member is on the specific location and select one of the following methods of communication:

(a) *Acoustic call*
(b) *Video call*
(c) *Uploading a photo*
(d) *Sending a message*

By selecting the *Download* option the user can save the map on his/her device, so that it will be available in case there is no network access at some point (offline use).

H. Thermal Camera Functionality

The options provided to the user by the Thermal camera feature of this mobile application (as presented on the screen of Fig. 11.7a) are:

- Related Information (name of the station, date, time)
- Positioning Spots
- Sending/Saving an Image

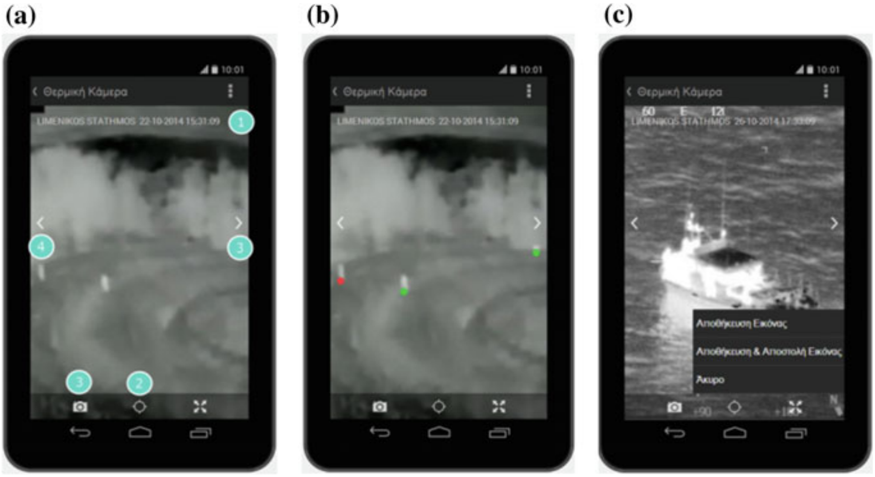

(a)
(b)
(c)
10:01
Θερμική Κάμερα
LIMENIKOS STATHMOS 22-10-2014 15:31:09
1
2
3
4
LIMENIKOS STATHMOS 26-10-2014 17:33:09
60
E
121
Αποθήκευση Εικόνας
Αποθήκευση & Αποστολή Εικόνας
Άκυρο
+90
+180
N

Fig. 11.7 **a** Thermal camera functions, **b** thermal camera—positioning spots, **c** thermal camera—sending/saving an image. Source http://iwapihex.xlx.pl/helicopter-with-thermal-image.php

- Slide shows with swipe moves (right/left).

By selecting Positioning Spots the user can see spots on locations where humans are present (Fig. 11.7b). Additionally, by selecting *Sending/Saving an Image* (Fig. 11.7c) the user can store or send the image from the thermal camera to the members of the operational team or the captain.

Mobile Application at the Central Operations Center

This smartphone application complements the previous ones as it may be used by members of the general staff of the Ministry of Shipping, which supervises all operations conducted in Greek marine waters. The main functions of this application include:

- Contacting the members of the operations (Regional Chiefs of Staff—Harbor Masters—Vessels' Captains)
- Viewing the map and AIS information
- Viewing the position of members of the operations
- Viewing image from a thermal camera
- Viewing image from a radar
- Viewing image from a camera mounted on a pc or a laptop
- Voice recording capability
- Viewing image/video gallery
- Sending/receiving messages

Using the map (Fig. 11.8) it is possible to view information relating to vessels operating in the area of interest. By selecting a ship icon (Fig. 11.8—point 1) a new window appears with information regarding the respective moving or anchored vessel.

I. Main Screen Functionality

   - Current Operations: View of the operations that are predetermined.
   - Current Operation: View of the current operation, which is created by the user.
   - Map—AIS: View of map—AIS data on the main screen.
   - Thermal Camera: View the thermal camera feed on the main screen.
   - Radar: View the radar feed on the main screen.
   - Camera: View on the main screen feed from a camera installed on the pc or laptop of a user.
   - Recording: Ability to record conversations among the user and other members of the operation (e.g. regional chief of staff, harbormaster, captain of a ship).
   - Video/Pictures: Viewing image and video gallery. With this option the user can be for example redirected to the Video/Image Collection ship external source screen (e.g. sample screen as a sole example from marinetraffic.com).

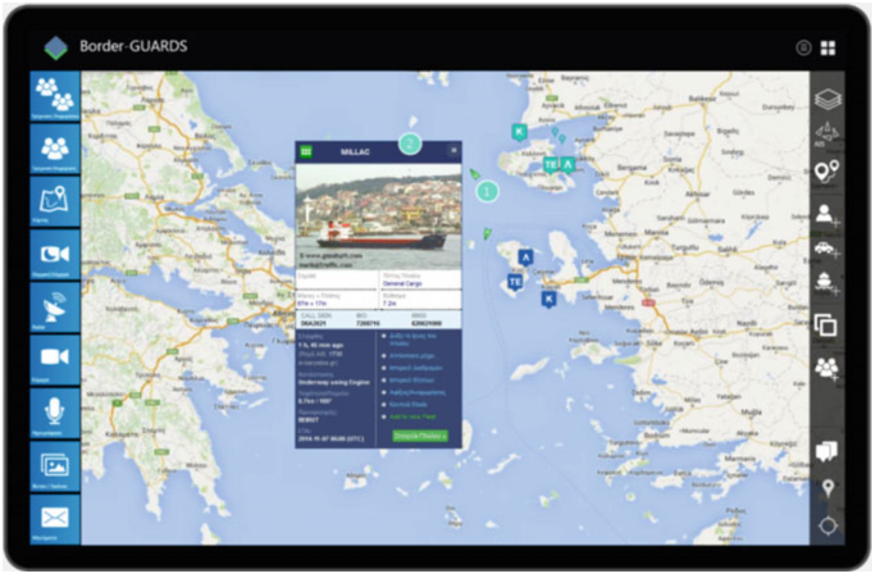


Fig. 11.8 Vessels' information

- Messages: Ability to send messages to members of the operation.
- Notifications: Showing notifications for incoming messages, pictures or videos.
- Messages' Window: Viewing a window with the latest incoming messages.
- Grid View: View of the map, thermal camera, radar and video camera on the main screen of the application as a grid view.

J. AIS Data Functionality

Information about the movements and vessels' positions is displayed on the map of the application (Fig. 11.8). This information is based on AIS data, navigation data, and information from ships and ports (Table 11.1).

Table 11.1 Indicative icons used by the application to distinguish users

Icon	User type—members of the operations
TE	Regional chiefs/operation center officers
Λ	Harbor officers
K	Vessels' captains/vehicle officers
📍	Members of the operational team

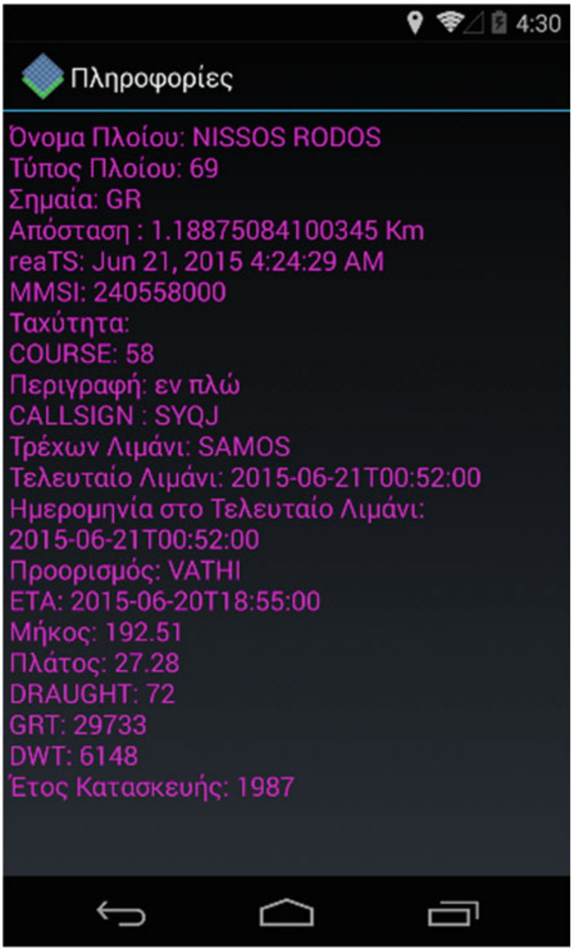
The settings for the displayed AIS information may be altered by selecting *AIS* (Fig. 11.8—numbered item 1), which includes the following options (Fig. 11.8—numbered item 2):

- Enable/Disable AIS
- Filters for the displayed information: for different types of ships e.g. passenger or cargo, moving or anchored, etc.

K. Outdoor Verification

The system has been tested in a real life environment using vessel and Radar at the Aegean sea in the wider region of Attica. The results have proven to be encouraging and promising. Detection of unidentified vessels has been performed using analysis of incoming streamed data from the radar, combined to AIS information and depicted on the mobile single screen platform (Fig. 11.9).

Fig. 11.9 AIS detailed information of a single ship on a smartphone in a lightweight text view to support low transmission and bandwidth network delivery



Όνομα Πλοίου: NISSOS RODOS
Τύπος Πλοίου: 69
Σημαία: GR
Απόσταση : 1.18875084100345 Km
reaTS: Jun 21, 2015 4:24:29 AM
MMSI: 240558000
Ταχύτητα:
COURSE: 58
Περιγραφή: εν πλώ
CALLSIGN : SYQJ
Τρέχων Λιμάνι: SAMOS
Τελευταίο Λιμάνι: 2015-06-21T00:52:00
Ημερομηνία στο Τελευταίο Λιμάνι: 2015-06-21T00:52:00
Προορισμός: VATHI
ETA: 2015-06-20T18:55:00
Μήκος: 192.51
Πλάτος: 27.28
DRAUGHT: 72
GRT: 29733
DWT: 6148
Έτος Κατασκευής: 1987

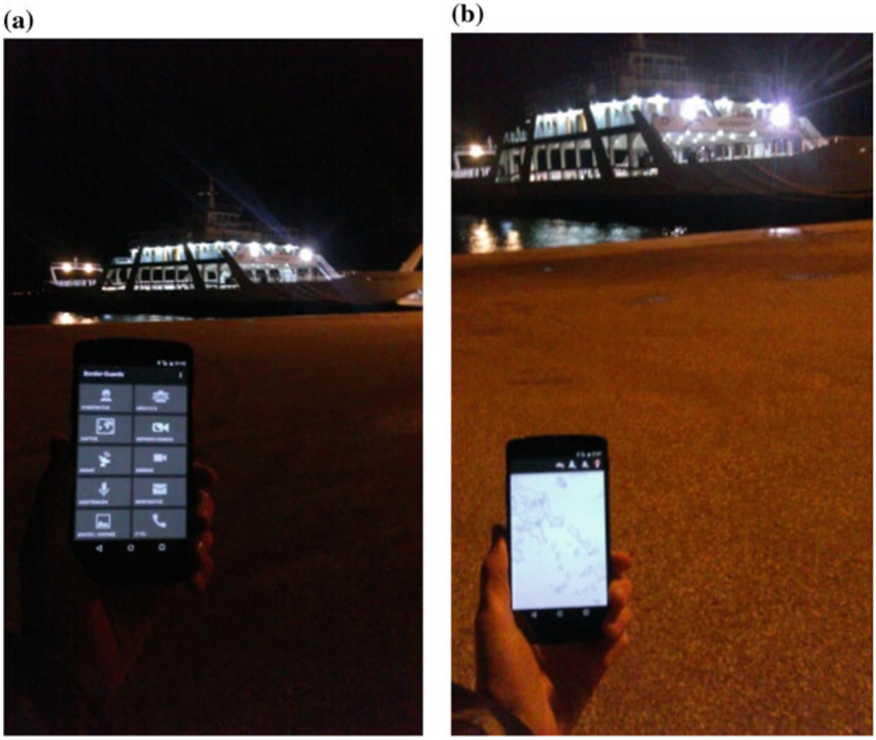


Fig. 11.10 **a** Inspection view by the ferryboat, **b** offline map location by the ferryboat (a boat used to ferry passengers, vehicles, or goods across open water, especially one that runs to a regular schedule)

Further testing for the communication platform has been done in Western Greece in the wider region of Patras where multiple harbours are available including one of the largest exporting gates of the country (Fig. 11.10).

Conclusions and Future Work

In this work, we have presented systematically a data fusion platform that allows for detection and tracking of moving objects for the case of marine surveillance. The proposed approach takes advantage of current communication and networking technologies for surveillance and builds upon them a single window data fusion environment to assist and enhance detection, early identification and tracking of moving targets for security and surveillance applications. Contrary to other approaches, the proposed solution contributes the same functionality both at the operation center and the end-user officer operating on a vessel at the sea or at the harbor on foot or on a vehicle. In this way we provide significantly high overall situational awareness to all the operational staff using the system beside the operation center. The data is appropriately transformed before delivered to officials at low bandwidth networks, therefore no full raw images are delivered if not absolutely necessary. Authorization mechanisms are enforced across the mobile systems, group and operational teams are supported as well as strong authentication. The solution is available on all possible form factors such as computers, laptops, smartphones, tablets and smartwatches. We have taken into account the exchange of real time data with any maritime/aerial asset, no matter where the operation is conducted and independently from the existing surveillance infrastructure.

The methods and technologies presented here may be transformed easily for the detection of marine pollution incidents. The proposed approach can accommodate further innovative designs of naval architecture and marine engineering, for platforms (watercrafts, small boats).

Acknowledgements This research has been partially co-financed by Synergasia Programme 2011–2015. This Synergasia Programme is co-funded by the European Regional Development Fund and Greek National resources (Project Title: border guards—Border Geomatics Utilities for Area Reconnaissance & Decision Support, Project Code: 11SYN_9_781).

References

1. National Single Window Prototype for administrative procedures. http://www.emsa.europa.eu/emsa-homepage/2-news-a-press-centre/news/2317-national-single-window-prototype-an-electronic-solution-for-simplifying-administrative-procedures.html
2. SafeSeaNet. http://www.emsa.europa.eu/ssn-main.html
3. AIS: Automatic Identification System. http://en.wikipedia.org/wiki/Automatic_Identification_System
4. Sakkopoulos E, Viennas E, Paschou M, Ioannou Z-M, Gkantouna V, Sourla E, Tzimas G, Sioutas S, Tsakalidis A (2015) Mobile data fusion from multiple tracking sensors to augment maritime safety: mobile detection, early identification and tracking of moving objects. In: IEEE international conference on mobile services 2015. NY, USA, pp 112–119
5. Kliem A, Hovestadt M, Kao O (2012) Security and communication architecture for networked medical devices in mobility-aware eHealth environments. In: IEEE first international conference on mobile services, MS 2012, 24–29 June, 2012. Honolulu, Hawaii, USA, 2012, pp 112–114
6. Prototyping, Software Prototyping. http://en.wikipedia.org/wiki/Software_prototyping
7. Lyndon C Design better and faster with rapid prototyping. http://www.smashingmagazine.com/2010/06/16/design-better-faster-with-rapid-prototyping/
8. Beaudouin-Lafon M, Mackay WE (2002) Prototyping development and tools. In: Jacko JA, Sears A (eds) Handbook of human-computer interaction. Lawrence Erlbaum Associates, New York, pp 1006–1031
9. Justinmind Prototyper. http://www.justinmind.com
10. Grubitzsch P, Schuster D (2014) Hosting and discovery of distributed mobile services in an XMPP cloud. In: IEEE third international conference on mobile services. Anchorage, AK, USA, 27 June–2 July 2014, pp 47–54

11. Ruta M, Scioscia F, Ieva S, Loseto G, Di Sciascio E (2012) Semantic Annotation of openstreetmap points of interest for mobile discovery and navigation. In: IEEE first international conference on mobile services, MS 2012, 24–29 June, 2012. Honolulu, Hawaii, USA, 2012, pp 33–39
12. Zhang X, Hu G, Duan N, Gao P, Dong W, Zhu W (2014) Scalable mobile data streaming with trajectory preserving partitioning. In: IEEE third international conference on mobile services. Anchorage, AK, USA, 27 June–2 July 2014, pp 16–23

Chapter 12
Mobile Stand-off and Stand-in Surveillance Against Biowarfare and Bioterrorism Agents

Manousos E. Kambouris

Abstract Engineered microorganisms, microorganisms traveling through massive human transportation systems to intercontinental distances and the natural processes for fast-track microorganism evolution, especially to counter antibiotics, secure the continuous presence of infectious diseases within the foreseeable future. In consequence, bioweaponeers, especially bioterrorists, will find excellent grounds for nefarious improvisations by exploiting novel agents with enhanced virulence characteristics, deliverable by various means but especially by spraying aerosolized forms. To counter the spread of such agents, along with the just as hazardous prospect of unintentional harmful agent release due to biotechnological accidents, more stringent biosurveillance schemes must be enacted, possibly 24/7, preferably integrating networking principles, state-of-the-art assets for both sensing and sampling applications and the use of inexpensive unmanned platforms, preferably mobile and even better moving in three dimensions (UAVs) so as to increase the reach, depth and persistence within surveyed space. The new tendency in post-sampling sensors will be prepackaged point-of-care assays with limited or no need of energy source and consumables, detecting 3-D structures or the nucleic acid signal for identifying agents so as to implement reactive, targeted countermeasures (decontamination, treatment). Proactive, general countermeasures, such as alert, sampling procedures and protective measures will depend on pre-sampling sensors, operating on spectroscopic principles and usually using UV-Laser Induced Fluorescence principle and carried onto manned and unmanned aerial and ground platforms designed for reconnaissance and surveillance with exchangeable payloads.

Keywords Bioagents · Inhalation · Ingestion · UAV · Stand-off
Stand-in · Sampling · PCR · Point-of-care · UV-LIF · Spectroscopy
Immunodiagnosis

M.E. Kambouris (✉)
Department of Pharmacy, University of Patras, Patras, Greece
e-mail: mekambouris@yahoo.com

M.E. Kambouris
Department of Food Technology, ATEI of Thessaly, Karditsa, Greece

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_12

Introduction

The millennia-old checkered relations between microbiota and the humanome (defined as the entirety of human subjects plus any directly dependent biological or non-biological entities) are, for some decades, at a razor's end: Infectious diseases did not become redundant and the optimism which arose with each new wave of important discoveries has been evidently misplaced [1]. Infectious diseases are here to stay, and prospects are bleak as to their use for bioattack/bioterrorism purposes [2], since multiple factors seem favorable: agents become more resistant, new infectious agents emerge and malefactors may acquire more effective, streamlined know-how and equipment [3–7]. The new eventuality of hybrid warfare, a neologism describing centuries-old practices [8], points towards a unified/joint surveillance, if not a unified/joint response system as well, which will encompass three fields: military and civilian healthcare, agriculture-plant growing protection and fishery & livestock breeding security [9]. The last two are becoming of more interest of late, as pathogens routinely break host species barriers, not only from animals to human (with Anthrax, HIV and Brucellosis coming to mind [10]) but from plants to human as well, with established plant pathogens, as are the fungal genera *Fusarium* and *Aspergillus*, causing severe conditions, especially, but not exclusively, by means of toxicoses. But this is just one aspect of the problem: human community welfare and survival is interdependent with the primary sector for quite a number of amenities which pertain the secondary and tertiary sectors as well. Thus, attacking crops and breed facilities may have dire consequences in social and political levels. The late blight of mid-19th century caused the depopulation of Ireland [11].

Common threats dictate pooling together surveillance and response means and methods as the only logical, economical and integrated way to capitalize on given resources. Particularities will surface, but this happens even within one of the affected fields—medicine.

Current Status and Projected Threat

The internal structure of humanome, as a distinct entity within the biosphere, is one of the most ominous factors. In engineering terms the humanome has expanded disproportionately, both in terms of rate and of mass when compared with other constituents of the biosphere. The social development of the human species does not delegate the most vital roles to the biologically fittest and most robust individuals in terms of natural selection, but of internal, cultural assignment of social status is followed. As a consequence, if some strategically placed but biologically unfit individuals are debilitated by an epidemic of even limited proportions, the system might well crumble and massive death toll may ensue, due rather to the systemic inability to sustain vital community functions—such as the sufficient and regular supply of necessities-than due to the lethality of the epidemic proper. Similarly, failure to safeguard the vital primary sector, an underguarded, expanded and vital target, can lead to famine and social implosion, much to the like of Ireland as mentioned above [11].

At the same time, ominous prospects arise from the expansion of proliferation of potentially hazardous biological resources. For the current study proliferation is defined as “the access to live cells, virions or to relevant genetic material augmented by the technical means and cognitive skills to productively use them into viable lifeforms”. Contrary to usual definitions which include malicious purpose, the motive is omitted herein, as motives change not only by the physical succession of operators, but also by handling and psychological manipulation of given operators and their chain of command.

To the above list on proliferation material, functional viroids and prions are likely near-future additions, although currently their weaponization potential seems negligible and they are practically never mentioned as possible biowarfare/bioterrorism agents.

Once current proliferation dynamics are taken into account [3], malicious potential, collectively and provisionally defined as “meta-infectiousness”—as it drastically enhances the parameters of natural infection events—seems to diverge to two distinct possibilities: Either ultra-evolved agents, engineered to perfection by high-tech, state or non-state developers operating advanced instrumentation, enjoying access to specialized know-how and supported by ample funds and dedicated infrastructure; or large numbers and medium quantities of low-tech and diverse agents developed and produced in dispersed, chaotic pattern and available to every moderately motivated, trained, equipped and funded player/actor; within this possibility one can count massive accidents, spontaneous or perpetrated, in biotechnology concerns. The latter cases, if considered collectively, will represent an indeed massive and diverse threat, but not particularly high-tech and thus generally detectable, traceable and treatable, at least in principle.

The Surveillance as a Constituent to Biodefense

To counter the meta-infectious threat, any conceivable strategy should be based on (1) prompt management of outbreaks, preferably proactively (suppression, containment-prevention, protection) but reactively as well (treatment, decontamination) if need be; (2) extended surveillance so as to implement the (1). Surveillance holds the key to successful management, as it allows warning and suitable countermeasures, both proactive (protection suites and biosafety protocols) and reactive (medical prophylaxis and treatment, decontamination) [9] (Figs. 12.1, 12.2).

Surveillance strategies can be differentiated by the frequency of coverage of the surveyed space/entity to intermittent/repeated as opposed to the persistent/continuous (see Figs. 12.3 and 12.4). The sensors used allow either “**stand-off**”

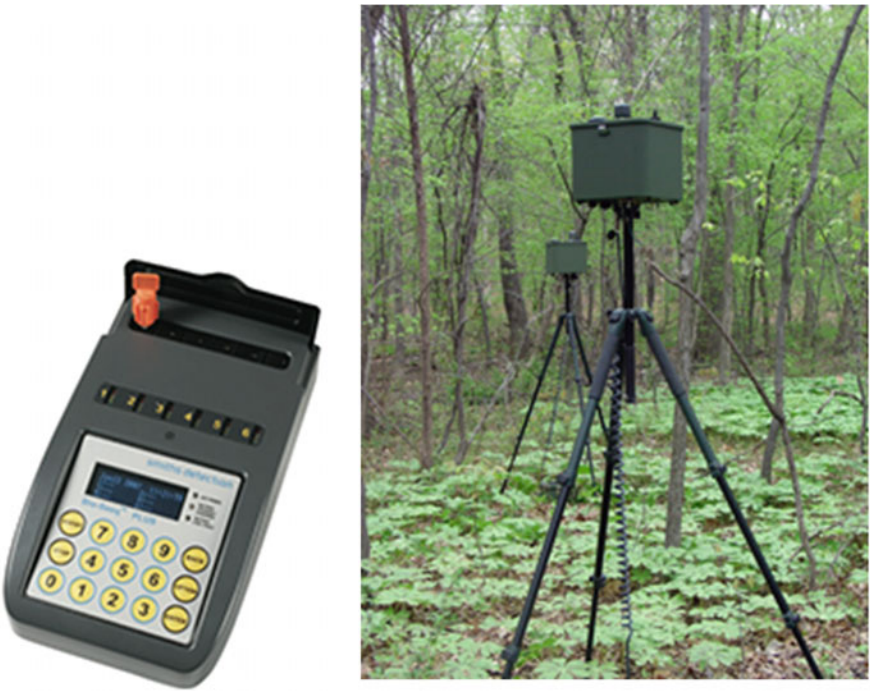


Fig. 12.1 **Bio-seeq®** hand-held, battery operated analyzer for in situ identification of bioagents-Left. Right, **BAWS®** fixed, movable sensors for surveillance in linear formation

Fig. 12.2 Long-range (or "High Altitude Long Endurance"—HALE) huge UAVs, like the US Q-4 family, may carry sensors and assaying equipment to process samples in situ and transmit results by data link while on patrol

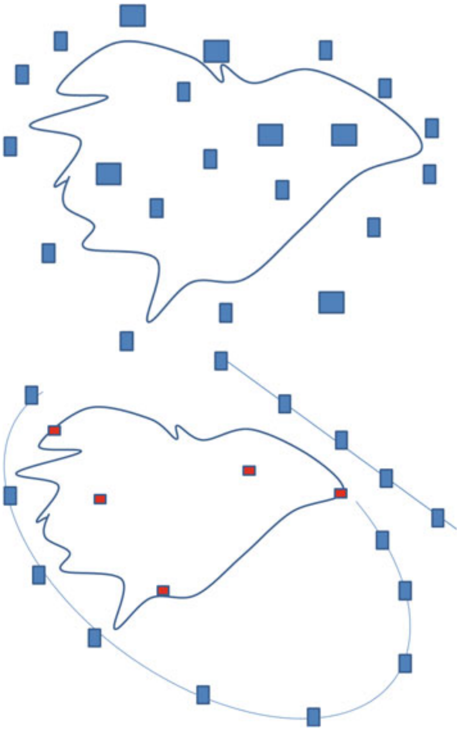


Fig. 12.3 In-depth surveillance pattern with fixed sensors/samplers (blue squares). Different linear surveillance patterns (straight or curved lines, open or closed lines) with fixed sensors/samplers (blue squares) compared to sensors/samplers delegated to vital localities (point surveillance—red squares)

detection, which generates results from a distance, by long range, "imaging sensors", or "**stand-in**"; the object of the latter is the analysis after sampling for detailed identification, insinuating contact or close proximity to the surveyed entity [9]. The combination of the sensors used with the surveyed entity/space produces three basic operating modes: **Overviewing** (long range, high-ground or other vantage point); **intrusive** (performed in confined and/or isolated spaces, irrespective of range but usually "stand-in"); and **distributed**. The third mode, distributed surveillance, uses sensors or samplers in multiple, dispersed loci which collect and download results and/or samples simultaneously or in any order ("in-" or "out-of-phase" respectively).

Surveyed entities are the environmental fluids (air, water masses), contact and edible objects; thus surveillance should be focused to air (both sampling and sensing approaches are applicable); to fresh and salt water (where sampling is applicable but remote sensing is not, at the time); and to edible/drinkable objects and contact environment, (where sampling is standard procedure as well, but sensing might be tricky at best). Very few agents found in water and contact environments can directly penetrate and cause invasive infections; on the contrary, airborne agents have direct and fast access (by inhalation) to the vital and rather vulnerable mucosa of the respiratory system, but also to other vulnerable sites (eye, ear and mouth mucosae) and constitute the most immediate threat [9].

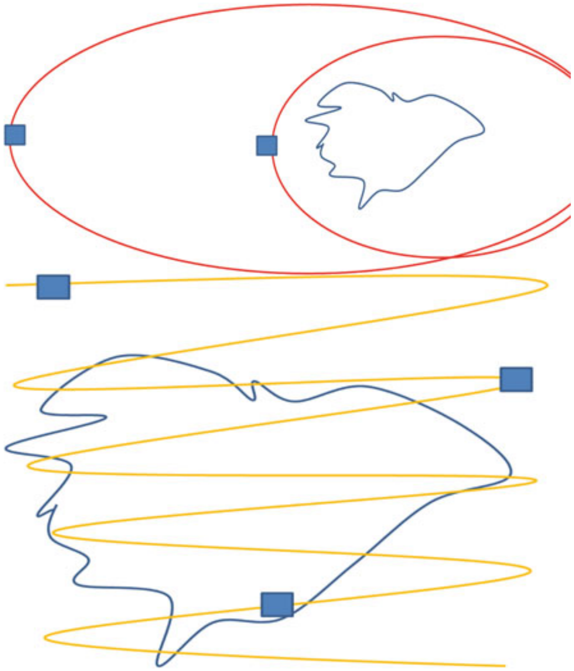


Fig. 12.4 Orbital mobile surveillance pattern in successively decaying ellipses, producing closed linear coverage; the geometry depends on the weather and the platform's velocity. The shape offsets the delay in moving from one side to the other so as not to miss an incoming cloud with a single platform. The surveyed area is covered by fewer sensing/sampling assets compared to in-depth formats

Moreover, for crop disease, especially in agroterrorism context, airborne contagion is the most ubiquitous mode of contagion, followed by vectors and pests [12]. Thus, the air is the main object of surveillance by both sensing and sampling assets.

The most basic, simple and thus cost-effective and flexible way to implement distributed surveillance is to dispatch mobile teams (ground- or airborne, mechanized or foot-mobile) for consecutive, in situ sampling and subsequent analysis. This approach highlights the need for portable, point-of-need assays, preferably hardware-independent [13] but, in times of need, plainly portable -or, even better, hand-held, battery-operated -and selectively networked instrumentation [9] fits the bill as well (Fig 12.1).

A new Surveillance Strategy, as proposed here, should seamlessly integrate the following aspects and attributes: short update cycles to keep the picture relevant; massive-in number and variety—but distributed-in spatiotemporal terms—functions (imaging, sampling and processing) compiled in fewer steps and levels compared to current approaches; ability to encompass all environments and all carrier fluids (air, soil, water, urban) and live populations of interest (human, animal, plant [11, 14]); scanning of increased volumes (not areas) and of confined, denied or inaccessible space by ad hoc dispatch of assets, thus reconfiguring the surveillance pattern and/or extending the surveyed volume as needed upon intelligence. The deployable assets needed, such as stand-off and stand-in sensors, and the respective processing facilities, might shrink in numbers or expand in coverage, according to necessity, by applications of drone technology [9]. Lastly, the extended surveillance footprint can be enhanced by robust telematics of standard, commercial amenities and costs, implementing near real-time surveillance [15] (see (Figs. 12.1 and 12.2).

Thus the wider concept can use current approaches; but for optimized results it should be focused, at least sensorwise, on two currently developing technologies: stand-off surveillance, such as staring UV and IR sensors for early warning and Novel Diagnostics for identification and classification.

Sensor Categories

Agent detection instruments are used in pre-sampling context and operate by acquiring the signature (IR, UV) or the size (or a combination of the two) of particles dispersed in the air, preferably in extremely high numbers [9] so as to form a cloud. Laser amenities, preferably of the LIDAR (Laser Imaging, Detection And Ranging) principle, such as the Short-Range Biological Stand-off Detection System/SR-BSDS [16] or simple laser spectroscopy in the UV (UltraViolet Aerodynamic Particle Sizer® spectrometer [17]) operating in the UV wavelength are the only way for remotely detecting live agents; the ceiling might have reached 30 km, the open-source operating range of the LR-BSDS. The UV illumination causes fluorescence to aminoacid residues containing aromatic rings (mainly to tryptophane, but also to tyrosine and to phenylalanine) and to the reduced form of nicotinamide adenine dinucleotide—NADH and to flavin compounds [9]. On the other hand, IR wavelengths allow detection, ranging and assessment of size and motion of a particle cloud (WindTracer® [18]). Some available systems combine the two functions into multi-wavelength formats (i.e. the SR-BSDS [16]). More detailed analysis, such as classification even to kingdom-of-life or domain-of-life level, is achievable only by stand-in techniques which require sampling [19], such as bioluminescence to establish vegetative, metabolically active status (as with the M31 BIDS [20]), preferably combined by particle counting/sizing (as with the Interim Biological Agent Detector [21]) in order to discriminate environmental inorganic particles, benign cells and threatening bioagents of different sizes (Prokaryotes vs Eukaryotes) and live cells versus spores.

The excessively specific tests, developed more as solutions seeking the right questions rather than on a need-to-have basis, answer nowadays with great accuracy and sensitivity very specific questions. The development of new methods, products and services is mainly attuned to profit and fame and less to robustness under unfavorable conditions, reduction of cost and integration into wider schemes, which would have forwarded standardization, commonality and redundancy. In many cases diagnosis may or even must be performed in this manner, but this is not the case with (Bio) surveillance, which should remain adaptable, affordable and well-defined in means and ends. Thus, a set of solutions originating from the military field come to attention as more suitable than purely diagnostic amenities.

At the turn of the century the White House declared the biothreat of outmost importance for the US security [22], and a quick reaction program was initiated, fully addressing most of the main aspects of the biothreat, as perceived less than a decade earlier. The main element of the American program was “adaptability” and relevant aims and milestones were released in 2006 for auditing [23], as they addressed issues and solutions throughout the spectrum of the evolving biothreat: enhancement of immune response by diverse approaches was instrumental, but other approaches were also developed as the immune response may be suppressed, compromised, diverted or de-tuned regarding its initiation, intensity, duration, specificity, and effectiveness [*ibid*].

But the most important lines of research, impacting surveillance and diagnosis, were the multi-factorial analyses of the biosignature of an infection, thus resolving diagnosis faster than with conventional isolation and study of the pathogen and circumventing possible precautions of the perpetrator to alter basic symptomatology and key detection/recognition entities and loci in order to lead to misidentification and erroneous treatment. The latter is one of the characteristics of the 3rd generation of bioagents, which are thoroughly engineered or outright artificial, in contrast to 1st generation agents (isolated, selected and cultured, as the ones used by the Japanese Unit 731 and the ones implicated in the Sverdlovsk incident in 1979 [24, 25]) and to 2nd generation (optimized but not engineered, as in the Anthrax letters incident in 2001 [26]).

Despite methodological breakthroughs, for infection control two sensing approaches/methods were, are and shall remain relevant: antiserum-based and PCR-based detection and identification [27]. MALDI-TOF prevails in identification-grade spectroscopy, without being particularly robust, as slight differences in the chemical composition of the sample lead to misidentification—thus its suitability for field use is debatable [28, 29].

The serological method is sensitive enough, robust in field conditions and easy to apply even without mechanical instrumentation, in energy- and hardware-independent assays, as are the pregnancy tests. Strips with fixed Antibodies perform one-off expendable tests applicable to any agent, including toxins and possibly prions [30], without any hardware or source of energy, a feature invaluable either in the field or in failed or Third World states [31]. Still, serological testing might be countered by altering the serological signature of an agent. New genetic, protein and organism engineering methodologies probably allow deletion or alteration of epitopes without impacting functionality, especially in cases where the epitope is different than the effector moiety/ies. In this manner, misidentification or even undetectability of the engineered agent may be achieved, especially against serological arrays using monoclonal antibodies [9]. With that in mind, the CB-47 [32], a US Pentagon program in advanced stage back in 2005, referred to an enhanced immunodiagnosis platform allowing a very prompt identification sequence which would cover the detection not only of cell-free molecular effectors, such as toxins, prions and viroids, but also of exposed and exposable cellular/viral effector molecules [33], expanding on previous platforms such as Guardian®, 4WARN® [5], Sensitive Membrane Antigen Rapid Test (SMART®) [34], with or without reading instrumentation. More recently, 3-D molecular recognition may be performed by non-immunomediated concepts as is the use of aptamers and of antimicrobial peptides [35].

But the big bet was always the nucleic acids assay technology, reserved for nucleic-acid—containing agents; practically it refers to all live agents with few exceptions. Up to now it needed expensive precision hardware with demanding infrastructure and resources: from the field-lab PCR, such as the APSIS® [36], RAPID®, and LightCycler® [37], the evolution passed through the handheld Bio-Seeq® [38] and finally the revolutionary CB-64 [23] of the US Pentagon emerged in the previous decade; it was to be a genomic, microarray-based platform with on-array amplification and sequencing technology to detect, identify, quantify and implicate genetically engineered biothreats. The available loci of the array and their fidelity to actual sequences were not an issue at the time, as the assembled genomes were the most advanced threat concepts, while the simply bioengineered agents were considered the main threat [*ibid.*]. Though, in the era of synthetic genomes [39] and microbiotes produced to specifications [40], this approach was deemed outdated and probably never fielded. Genetic engineering and alternate coding (based on the inherent degradation of the genetic code) can be used to fool such detection systems, but they are costly, difficult and laborious. Random priming, real-time protocols and multi-locus approaches can overcome such countermeasures expeditiously.

On the other hand, DNA analysis is less easy and user-friendly as it requires more steps, but much more robust, since both whole organisms and single effector moieties depend on DNA sequences to function (the former) or to be produced at any cellular level (the latter), with the possible exception of prions and some RNA viruses. The role of this type of analysis might be fulfilled by novel approaches, which keep costs down, do not require instrumentation, offer inherent adaptability to evolving threats and are user-friendly. Point-of-need concepts, using isothermal amplification, with Recombinase-Polymerase Amplification being the paradigm [13], may revolutionize on-the-spot analysis by dispatched teams which survey ad hoc or routinely, a concept-and a technological approach—not altogether novel [41]. Energy-independent assays in ready-to-use platforms (chip/strip assays [29]), employing chemicals prepackaged in capsules, or pellets; lateral flow matrices; colorimetric testing in visual colors without the need of readers and exciters; all further the battery-operated, current state-of-the-art diagnostics, like the Bio-Seeq® instrument (Fig. 12.1 left [38]) towards a much more flexible, adaptable and affordable concept.

The problem with DNA and nucleic acids detection in general is that toxins, a well-known public health issue and a bioweapon *par excellance* [42] cannot be detected at all, if used in a cell-independent mode, i.e. as a (bio)chemical weapon.

The Spatial Aspect of Surveillance

Mobile ground-, air-, or waterborne platforms, manned or unmanned (UGVs, UAVs, USVs, UUVs respectively) are able to create dynamic, deployable and responsive surveillance networks. Such networks can substitute, replace or simply augment fixed ones, the latter based on fixed or movable/semi-fixed sensors (such as BAWS$^{®}$, Fig. 12.1 right, [9]). Thus, mobile platforms provide a dynamic component in an originally static layout usually operating on a deterministic logic, covering most vulnerable, exposed or compulsive spots. Moreover, at least in principle, there is an economic argument as well: moving assets cover more space with fewer units and their distribution can adapt to evolving situations (Figs 12.3 and 12.4).

Airborne assets in particular (Fig. 12.2) are ideally suited to survey, detect, locate and sample allegedly biohazardous material. The speed and the three-dimensional mobility of airborne assets, compared to other approaches, allows an extended surveyed space, creation of depth at the surveyed area and fast redeployment of assets if need be, thus permitting contingency planning. Additionally, the selection of either rotary or fixed-wing platforms allows different combinations of accessibility and deployability on one hand and speed of transition, autonomy and range on the other. The motion through air allows more permissive moving environment and better navigation, saving transit time and allowing extended loitering in both spatial and temporal terms.

There are different levels of sophistication in the application of the above, impacting on total cost, effectiveness and reliability: The simplest format is to install, as an add-on, a biocollector on a small helicopter or tactical or security UAVs [43]. It can be mounted ad hoc, as external load or semi-permanently within a conformal “bulge” attached directly on the airframe, preferably the fuselage. The sampler may be powered either by the generator/battery of the carrier platform, or through exploiting Bernoulli’s principle to ingest airstream and particles. For the latter, a system of airducts through which air is sampled is required, and it can be improvised by rather straightforward modifications in airframe or pod design. Rather inexpensive aerial platforms carrying basic, slightly modified typical air samplers and low-cost GPS satisfy fully this approach. The collected sample(s) are processed off-board, when the platform concludes its mission, at a base facility equipped with all necessary diagnostic hardware for field or lab operating conditions. Sampling proper, when considering unmanned platforms, may be performed either by remote control or in predefined mission parameters; manned platforms are more responsive and human crews can exercise judgment to optimize mission-related events. Within a platform sortie (a cycle defined from the departure of the platform to its return to base, the same or different than the one of origin) sampling might be either unique-event, to check a perceived “hot” spot, or multi-event (more than one samples collected in spatially diverse areas)—depending on the autonomy of both the platform and of the equipment. Multi-event sampling sorties result in a more cohesive picture as they cover more space in limited time, minimizing the time-dependent drift.

The precise spot and the kinetics of the sampling are both of paramount importance for the success of the sampling operation. They can be determined autonomously by the platform if GPS-based or GPS-enhanced navigation is employed. Moreover, spatiotemporal precision in sampling is instrumental in creating an accurate, relevant, integrated picture by assembling the sampling data and fusing them with meteorology and external intelligence, such as satellite imagery or human—derived intelligence.

The high-end solution is much more elaborate, effective and expensive: it involves the installation of miniaturized and repackaged integrated diagnostic hardware onto the aerial platform, in either permanent, dedicated formats, or in exchangeable mission modules. Bigger, spacier platforms, such as endurance UAVs (like the Q-4 series, Fig. 12.2) [44] or small airplanes (like the C-12 series) [45] have enough room, lift, interfaces and provisions to carry and operate autonomously miniaturized versions of fully automated identification hardware. The provisions/interfaces include electrical power, cooling and adequate quantities of the consumables needed to perform repeated series of predefined tests, plus proper communications equipment to transmit the results of the on-board tests. Such hardware has been proposed for fielding since mid ‘00s in fixed or removable, unmanned, stand-alone sensors systems designed to survey positions of interest, almost unattended over extended times, with the option of being wired or wirelessly networked to create sentinel grids [9, 30]. But the Total Lab automation approaches which have emerged for routine lab applications [46] can be downsized for eventual inclusion in modular, palletized loads for aerial platforms. Current Total Lab formats can be used in strategic contexts, carried in large transport aircraft of C-130 size or bigger, for long—range deployments, while routine use in low operating cost platforms might be available within a decade.

The overwhelming advantage with automated onboard sample processing is its almost instant implementation, with minimal degradation of samples and the results being promptly uploaded to the surveillance network in near-real-time by even austere data links, allowing more time to report, assimilate, decide and react. Transmission of data output from diagnostic equipment should be achievable by quite a margin for datalinks used for exchanging high volumes of imagery, most of it of high-definition and in some cases of multi-sensor or even of multi-spectral origin.

There is of course a third solution, less ambitious and costly in development and purchase of equipment: to perform limited sample processing on-board while the actual data generation is implemented off-board, upon conclusion of the sortie. In such a concept, the direct involvement of expert personnel in the data generation steps (identification) allows for adaptability of the routines used to unforeseen challenges. A limited degree of in situ sample processing in some cases, especially in air samplers, contributes to increased autonomy by averting spontaneous sample degradation and reducing the volume and resources needed for sample storage. Consequently, more sampling events per sortie can be achieved, resulting in higher-volume data generation off-board.

The stand-off sensing approach refers to a fixed platform (such as SR-BSDS, [9]), or a moving asset, carrying a sensor able to detect an agent and perhaps classify or quantify it. Such items are particle counters and IR and UV sensors deployed at least since the '00s [9, 29]. Such platforms are basically devoid of means to collect samples, far less to perform identification and diagnostics, but combined stations may incorporate any number of assets, from sensors via samplers/collectors to identifiers [9]. Even if restricted to remote sensing only, they may transmit the sensor data in real time, so as to initiate if need be a follow-on mission for sampling/identification as early as possible, thus curtailing the total cost of implementing surveillance. This two-stage approach, with each stage performed by a different platform (1st surveillance/detection, 2nd sampling/identification), offers substantial savings in peace-time and internal security operations, given that the stand-off surveillance must be performed constantly and perhaps on a 24/7 basis, whereas stand-in surveillance/sampling only when and where an alert algorithm is activated [9, 19, 30].

The mobility of an aerial platform means that one or a limited number of sensors can survey the same area as an extended network of fixed sensors (Figs. 12.3, 12.4) and only if needed. If geometry and geography are permissive, the lapse in time needed for the transition of the platform to the adverse edges of the orbit might not cause a failure, especially if stand-off sensing is concerned (Fig. 12.4). And it is not just the reduced number of sensors needed that creates savings of scale: the platform may be used into other roles and be fitted with the biosurveillance/ biocollection payload in its standard bay/interface only when such a need arises.

Conclusions

In field assignments the tendency towards Virtual Presence (by links, communications, telematics, nets), which was invented to curtail the cost of expertise while retaining responsiveness to the point-of-need, might well be supplanted in Public Health and environmental protection by Remote Presence and Distributed Presence.

The former will be implemented through few, dedicated, strategically located, fixed, unmanned networked sensors [9], as intended for the last 10–15 years; but it shall be augmented by Unmanned mobile platforms (UAVs/UGVs). These allow great flexibility and cost-effectiveness in covering extended or confined space. The indices of the feasibility of the concept will be:

(i) the development of decontamination protocols for the dispatched robotic/remotely operated assets, in order not to bring back nor to further disperse and disseminate an agent;

(ii) the development of standardized, integrated "mission modules" with diverse sensors (preferably operating in multiple spectra and on different principles) and samplers in integrated assemblies, ready for loading on and off airborne and groundborne platforms. For a specific volume and weight, standardized interfaces and resources will allow plug-into the carrier, with different options and capabilities fitting in different sizes of carriers. Arrangement of current concepts of lab automation into volumes and geometry suitable for loading to small aircraft or big UAVs will decide the achievability of out-of-area, 24/7 surveillance. This development will allow full exploitation of the design of current manned and unmanned platforms with standardized payload bays and interfaces, exemplified by but not restricted to the light AHRLAC® [47] or even the massive U-2S [48] on one hand and Penguin® [49] and BAT-4® [50] on the other; a concept which increases flexibility versatility and lowers total costs.

In a different context, Distributed Presence refers to multiple, possibly isolated surveillance/intervention assets which contribute to a common picture or action. The idea is to have prepositioned, decentralized assets compiling a precise background picture so as to facilitate surveillance. In this case, the low cost of equipment, both diagnostic and utilitarian (communications, transportation) is focal to implementing the scheme. A straightforward approach is the ad hoc insertion of investigative teams for any amount of time, preferably encumbered with only minimal hardware but equipped amply with resource—independent, point-of-care assays [13]—a quantum—leap when compared to battery-operated devices, which were the cutting edge of technology at the turn of the century (Bio-Seeq®, Fig. 12.1 left [38, 51])—and might well still be. In Distributed Presence, the human factor remains supreme, although not mandatorily (fixed, networked sensors like the above mentioned-Fig. 12.1 right- BAWS® can do the trick [9]), nor exclusively: small, portable drones [52], modified so as to sample-or even sense—will be of importance to augment the reach and to reduce the risk of such dispatches.

References

1. Stuart-Harris C (1984) Prospects for the eradication of infectious diseases. Rev Infect Dis 6 (3):405–411
2. Bartholomew RA, Ozanich RM, Arce JS et al (2017) Evaluation of immunoassays and general biological indicator tests for field screening of Bacillus anthracis and Ricin. Health Secur 15(1):81–96
3. Frinking E, Sweijs T, Sinning P et al (2016) The increasing threat of biological weapons. The Hague Center for Strategic Studies
4. Inglesby TV, O’Toole T, Henderson DA (2000) Preventing the use of biological weapons: improving response should prevention fail. Clin Infect Dis 30(6):926–929
5. Švabenska E (2012) Systems for detection and identification of biological aerosols. (Rev). Def Sci J, [S.l.] 62(6):404–411
6. Morens DM, Fauci AS (2013) Emerging infectious diseases: threats to human health and global stability. PLoS Pathog 9(7):e1003467
7. English BK, Gaur AH (2010) The use and abuse of antibiotics and the development of antibiotic resistance. Adv Exp Med Biol 659:73–82
8. Murray W, Mansoor PR (2012) Hybrid warfare: fighting complex opponents from the ancient world to the present. Cambridge University Press, Cambridge
9. Primmerman CA (2000) Detection of biological agents. Linc Lab J 12(1):3–32

10. Al-Sadi AM (2017) Impact of plant diseases on human health. Int J Nutr Pharmacol Neurol Dis 7:21–22
11. Khalil AT, Shinwari ZK (2014) Threats of agricultural bioterrorism to an agro dependent economy; what should be done? J Bioterror Biodef 5(1):127–134
12. Carus WS The history of biological weapons use: what we know and what we don’t. Health Secur 13(4):219–255
13. Mayboroda O, Benito AG, del Rio JS et al (2015) Isothermal solid-phase amplification system for detection of Yersinia pestis. Anal Bioanal Chem
14. American Phytopathological Society ad hoc committee on Crop Bioterrorism Crop Biosecurity and Countering Agricultural Bioterrorism: Responses of The American Phytopathological Society APS Feature-2002-10 (2002)
15. Kambouris ME, Kantzanou M, Papathanasiou BA et al (2016) Towards total biothreat preparedness: expanded surveillance, joint monitoring, pooled resources and the genomic option. IJISET 3(7):384–389
16. Suliga W, Burnham RL, Deely T et al (1999) Short-range biological standoff detection system (SR-BSDS). Proc SPIE 3855:72–81
17. Huffman JA, Treutlein B, Posch U (2010) Fluorescent biological aerosol particle concentrations and size distributions measured with an Ultraviolet Aerodynamic Particle Sizer (UV-APS) in Central Europe. Atmos Chem Phys 10:3215–3233
18. Unknown, Windtracer. http://www.lockheedmartin.com/us/products/windtracer.html
19. Unknown, FLIR IBAC-2 Bio-Threat Detection & Collection. http://www.flir.com/uploadedFiles/flirGS/Threat_Detection/Biological_Detection/Products/Fido_B2/DatasheetFidoB2009012015EN.pdf
20. Unknown, Contamination Avoidance Programs, Annex A, p A-3. http://www.au.af.mil/au/awc/awcgate/nbc97/97anexa.pdf
21. Unknown, Contamination Avoidance Programs, Annex A, p A-4. http://www.au.af.mil/au/awc/awcgate/nbc97/97anexa.pdf
22. Unknown, Office of Press Secretary, The White House: 28 Apr 2004
23. Department of Defense Chemical and Biological Defense Program Annual Report to congress 2006:F33. https://fas.org/irp/threat/cbdp2006.pdf
24. Keiichi T (2005) Unit 731 and the Japanese Imperial Army’s biological warfare program. Asian Pac J/Japan Focus 3(11). http://apjjf.org/-Tsuneishi-Keiichi/2194/article.pdf
25. Sahla JW, Pearsona T, Okinakaa R et al (2016) A Bacillus anthracis Genome sequence from the Sverdlovsk 1979 Autopsy Specimens. *mBio* 7(5):e01501-16
26. Meek JG (2008) FBI was told to blame Anthrax scare on Al Qaeda by White House officials. New York Dailynews 2008. http://www.nydailynews.com/news/world/fbi-told-blame-anthrax-scare-al-qaeda-white-house-officials-article-1.312733
27. Broussard LA (2001) Biological agents: weapons of warfare and bioterrorism. Molec Diagn 6 (4):323–333
28. Unknown, MALDI Biotyper Poster hall 2016, Bruker: https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/Separations_MassSpectromtry/Literature/Brochures/1842996_Posterhall_MALDI_Biotyper_2016_ebook.pdf
29. Ludovici GM, Gabbarini V, Cenciarelli O et al (2015) A review of techniques for the detection of biological warfare agents. Def S&T Tech Bull 8(1):17–26
30. Pak T (2008) A wireless remote biosensor for the detection of biological agents. NNIN REU Research Accomplishments, 24–25
31. Hjelle B, Jenison S, Torrez-Martinez N et al (1995) Rapid and specific detection of Sin Nombre virus antibodies in patients with hantavirus pulmonary syndrome by a strip immunoblot assay suitable for field diagnosis. J Clin Microbiol 35(3):600–608
32. Unknown, DoD CBRN Defense Program FY 2003-5 Performance Plan 2004, p 95. http://www.acq.osd.mil/cp. Accessed 10 Sep 2015
33. Ding B (2010) Viroids: self-replicating, mobile, and fast-evolving noncoding regulatory RNAs. Wiley Interdiscip Rev RNA 1(3):362–375
34. Unknown, SMART-II Cholera O-1. http://www.nhdiag.com/cholera_bt.shtml

35. Sapsforda KE, Bradburneb C, Delehantyb JB, Medintzb IL (2008) Sensors for detecting biological agents. Mater Today 11(3):38–49
36. Unknown, Biological Warfare Agent Detection—APSIS, Bruker Daltonics Inc. https://www.copybook.com/companies/bruker-daltonics-inc/articles/biological-warfare-agent-detection-apsis
37. Christensen DR, Hartman LJ, Loveless BM et al (2006) Detection of biological threat agents by real-time PCR: comparison of assay performance on the R.A.P.I.D., the LightCycler, and the Smart Cycler platforms. Clin Chem 52(1):141–145
38. Unknown. http://www.securmar.com/detection/datasheet/trace/BioSeeq_Plus.pdf
39. Hutchison CA III, Chuang R.-Y, Noskov VN et al (2016) Design and synthesis of a minimal bacterial genome. Science 351, aad6253/1- aad6253/11
40. Gibson DG, Glass JI, Lartigue C et al (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329:52–56
41. Unknown, Department of Defense Chemical-Biological Defense Program Annual Report to Congress 2006, Annex B B-6
42. Anderson PD (2012) Bioterrorism: toxins as weapons. J Pharm Pract 25(2):121–129
43. Grivas K, Velegraki A, Kambouris ME (2008) Mid-Term deployability and geointegration concerns in biodefense sampling and detection hardware design and procedures. Defensor Pacis 22:111–116
44. Unknown, Global hawk. http://www.northropgrumman.com/Capabilities/GlobalHawk/Pages/default.aspx
45. Drew J (2015) Beale AFB farewells MC-12 as spy plane moves to Army and SOCOM Flightglobal. https://www.flightglobal.com/news/articles/beale-afb-farewells-mc-12-as-spy-plane-moves-to-army-417153/
46. Iversen J, Stendal G, Gerdes CM, Meyer CH, Andersen CØ, Frimodt-Møller N (2016) Comparative evaluation of inoculation of urine samples with the Copan WASP and BD Kiestra^{TM} InoqulA^{TM} Instruments. J Clin Mcrobiol 54:328–332
47. Sweetman B (2014) South Africa's Ahrlac Is Airborne, Ares. http://aviationweek.com/blog/south-africas-ahrlac-airborne
48. Pocock C (2015) U-2 Pioneers Open-mission Systems for U.S. Air Force, Ainonline. http://www.ainonline.com/aviation-news/defense/2015-09-06/u-2-pioneers-open-mission-systems-us-air-force
49. Unknown, Penguin C UAS. http://www.uavfactory.com/product/74
50. Unknown, BAT 4 UAV. http://martinuav.com/uav-products/bat-4/
51. O'Connell KP, Anderson PE, Valdes JJ, Bucher JR (2005) Testing of the Bio-Seeq (Smiths detection handheld PCR instrument): Sensitivity, specificity, and effect of interferents on *Yersinia pestis* assay performance. Edgewood Chemical Biological Center, 2005: ECBC-TR-437
52. Rosenberg Z (2013) Five companies win US Army contract for handheld UAVs, Flightglobal. https://www.flightglobal.com/news/articles/five-companies-win-us-army-contract-for-handheld-uavs-380816/

Chapter 13
Social Networks for Surveillance and Security: 'Using Online Techniques to Make Something Happen in the Real or Cyber World'

Ben Harbisher

Abstract This chapter examines the use of Social Networks for Surveillance and Security in relation to the deployment of intelligence resources in the UK. The chapter documents the rise of Military Intelligence agencies during both World Wars (such as GCHQ and MI5), and the subsequent use of these institutions to maintain order during peacetime. In addition to the way in which military organisations have used clandestine techniques such as double agents, spies, and various programmes designed for conducting Signals Intelligence, the Chapter offers an insight into how contemporary modes of communication (via mobile devices and the internet), shape the way in which intelligence agencies now gather information. The chapter also considers how the UK's intelligence community responds to National Security issues such as international terror attacks, and how additional threats such as political subversion are framed in National Security discourse as being the legitimising factors behind mass surveillance. Thereafter, the chapter examines how online techniques are used by Britain's intelligence agencies to maintain National Security, and how counter-intelligence strategies are being turned against the population to encourage political compliance. The chapter examines how online espionage techniques for entrapment, coercion, and misdirection, are being used to make something happen in the real or digital world.

Introduction

One of the most prominent threats posed to civilian populations and to the modern nation state today, can be considered in terms of attacks being conducted through digital worldwide networks. As opposed to traditional forms of munitions-based warfare and organised crimes, the unprecedented growth of digital communication

B. Harbisher (✉)
De Montfort University, The Gateway, Leicester, UK
e-mail: ben.harbisher@dmu.ac.uk

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_13

platforms and social media has posed a number of challenges to security providers and to private citizens alike. Modern risks presented by the widespread use of digital networks includes hacking and data/identity theft and other forms of cybercrime; the relatively new vogue in cyber warfare (such as techniques to disable the communication systems of hostile or unstable regimes); and in terms of espionage, the manipulation of online content for the purposes of disseminating propaganda and disinformation to influence the behaviour of specified targets.

The purpose of this chapter is to examine the conspicuous alignment of military and civilian programmes in surveillance, in relation to the securitization of social media networks and cyberspace. The chapter investigates a range of Signals Intelligence (SIGINT) protocols and the institutions who implement them, and explores the deployment of SIGINT operations against non-military targets such as the perpetrators of serious organised crimes and political activists in the UK. The chapter considers how online social networks have become an invaluable tool for maintaining national security, focusing on the agencies who conduct both mass and targeted forms of surveillance through this particular medium. The purpose of this research is to define which modern agencies are responsible for providing Signals Intelligence, and the uses to which this material is put. Here it is the aim to discuss two SIGINT strands in particular, by focusing on Britain's Counter-Intelligence and Effects capabilities. In this respect, Counter-Intelligence Programmes (known as COINTELPROs) exploit Signals Intelligence for the purposes of conducting covert actions against strategically valuable targets to the UK. The SIGINT capabilities of COINTEL agencies are defined by the way in which modern communications systems can be infiltrated and exploited to disrupt the activities of both terrorists and criminal organisations. In terms of Operational Effects, the intention of this subterfuge is to employ a range of online techniques to have an impact on recipients in the real world—ideally before any police coercion or military interventions are required. Security agencies today are able to infiltrate private chat rooms to monitor discussions, post online content to dissuade deviant actors from either planning or enacting their crimes, and even plant materials to discredit the reputation of key suspects during an operation.

However, to problematize the issue of conducting online surveillance to regulate social networks, the techniques used during SIGINT and COINTELPRO operations have in recent years been employed to subvert legitimate social movements in the UK. Evidence from a range of sources suggests that SIGINT and COMINT (Communications Intelligence) resources have been used to discredit non-military targets as well as volatile overseas regimes. Indeed such protocols have allegedly been used for maintaining public order in addition to fighting serious organized crimes and terrorism, that is, by gathering intelligence of public demonstrations and disrupting the communications systems of those involved [1]. In terms of open democracy, these clandestine practices pose a very serious threat to both civil liberties and human rights in the UK, just as much as they contravene the legitimate right of citizens to conduct peaceful demonstrations. A bigger question then, transpires in relation to the ultimate objective of such programmes. On the one hand, COINTELPRO activities may seek to pre-empt costly public order operations, yet, in another respect, does the state have the right to modify the behaviour or the opinions of activists—especially when such organisations may seek to campaign against unpopular or unfair policy decisions? What really it boils down to is a question of framing. In other words, of how intelligence organisations depict legitimate acts of dissent as threats to national security, and what the response of such agencies might be to these risks.

Code Breakers and Spies—From National Defence to National Security

The UK's National Intelligence Machinery can be divided into three categories, and the historical emergence of these organisations is important to observe. Principally, they can be defined as Military Intelligence, National Security agencies, and the traditional police authorities of England and Wales. Arguably, in today's intelligence market, these three institutions (and their many subdivisions) often share intelligence resources and a number of operational procedures for keeping citizens safe from harm. Government Communication Headquarters (GCHQ) for example, is responsible for gathering Signals Intelligence, and for disseminating this data to relevant parties. SIS/MI6 is the UK's Secret Intelligence Service and generally operates overseas to monitor persons or regimes of interest. Comparatively, MI5 is the UK's National Security Service and deals ostensibly with domestic matters (i.e.; internal threats posed to Parliamentary Democracy via subversion, by terror attacks, and from other forms of serious organised crime). Lastly, HM Constabulary represents a full-time permanent police force that patrols Britain's streets to prevent or detect crimes, and is otherwise used to maintain public order. Nevertheless, all of the above now works in unison in response to serious threats posed to National Security by terrorist organisations, and from designated groups intending to commit terror-type attacks against the UK [2: 4].

With the exception of HM Constabulary (whose legislative origin dates back to the *Metropolitan Police Act 1829*), the other institutions mentioned above are relative newcomers to matters of National Security. During the Nineteenth Century for example, the *Metropolitan Police Act* effectively formalised a number of different provincial watch services in the London area for the purpose of creating one unified police force. By 1839 the majority of independent watch services (such as the river police, mounted and pedestrian patrols), were unified to form the Metropolitan Police Force. Elsewhere in the UK, local parish and municipal policing services were formalised by the late 1860s, and were organised into local and regional divisions—with the exception of the Metropolitan Police Service and the City of London Police. Today, the Metropolitan Police Service manages the greater London area, and the City of London Police patrols London's financial district. The rest of HM Constabulary is broken down into regional administrative divisions and into the subsequent law enforcement agencies they manage in each district or town.

In terms of Military Intelligence, the majority of agencies now used to conduct espionage, surveillance, and counter-intelligence activities in the UK, began life during the early Twentieth Century. In 1909 the Secret Service Bureau commenced operations following requests made by the Admiralty and the War Office to monitor the Imperial German Navy (which was considered a substantial threat to the British Empire). According to military historian Christopher Andrew, the 'Secret Service Bureau got off to a confused start', with the appointment of two sufficiently experienced officers in the field of navel and military intelligence—neither of whom had been given much of an explanation as to what their particular roles might involve [3: 25]. As a result of this administrative gaffe, Captain Vernon Kell and Commander Mansfield Cumming decided between them, that it would be best to divide their responsibilities according to foreign and domestic intelligence affairs. Kell would gain oversight of naval and military intelligence at home, whereas Cumming's portfolio would govern equivalent operations overseas. It was this division of roles that later compelled the Secret Service Bureau to become two separate organisations.

As early as 1910, the distinction between the two intelligence agencies was already apparent in Whitehall, but it was agreed that both offices would continue to operate as the Secret Service Bureau. By the outbreak of World War One, the Bureau had formalised its overseas operations under the codename of the Special Intelligence Service (SIS), although it was officially recognised as part of the Directorate of Military Operations, MI1(c). SIS was responsible for establishing a network of spies either within the borders of Imperial Germany (which it did with limited success), or more effectively in the German Empire's neighbouring countries of Belgium, Russia, and France [4]. The domestic activities of the Bureau, however, had a lot more impact during this formative period—which was to conduct counter-espionage activities in the UK. Kell's group of operatives was commissioned to work under Section 5 of the Directorate of Military Operations group MO5(g), which, following the reconfiguration of the Directorate in 1916, became unofficially known as MI5 under the Directorate of Military Intelligence.

The MO5(g) was formally established by the Secret Service Bureau during this period, and by 1914 it was already conducting counter-intelligence operations against German spies on British soil. MI5's early accomplishments came from a pre-emptive strike against suspected foreign agents just as the Declaration of War was being announced. Working with MO5(g), Special Branch officers from Scotland Yard arrested 22 German spies as soon as the declaration was made on August 4, 1914. The Bureau later claimed that this nationwide campaign had removed the majority of German spies from British soil, and had given the UK the strategic edge in the forthcoming conflict. By 1917, MI5 (as it had become known) had accumulated a list of all foreign aliens living or working in the UK, and had processed the entrants according to the military risk they posed [3: 58]. At this time MI5's main resource (known as the Central Registry) contained over 27,000 records of individuals who the organisation held under suspicion for one reason or another.

This colossal archive had been assembled in-part from communication interception orders through which the postal mail of suspected foreign agents had been seized and then scrutinized. Although officially disputed by the Post Office, these interception orders were upheld by former Home Secretary, Winston Churchill by way of Home Office Warrants to confiscate any incriminating evidence (*Ibid*: 37).

As a result of its initial successes MI5 ventured into more political territory during the Great War, beyond its initial role as the UK's foremost counter-espionage organisation. In addition to hunting for foreign spies who were working in the UK, MI5 expanded its activities to conduct investigations into anything that might hamper the war effort. This included anti-conscription and pacifism groups who opposed the war, and also (as a means to occupy its expanding body of agents) to monitor Britain's Trade Labour movement as well. It was feared that industrial unrest provoked by the trade unions would cause the manufacture of weapons and munitions to cease, and that such organisations were vulnerable to infiltration by foreign agents who would incite dissent. However, despite its initial gains, following the First World War MI5 was subjected to nearly twenty years of executive scrutiny, administrative oversight, staffing and funding cuts, by consecutive peacetime governments. Although MI5 had flourished during the Great War, the scale of its operations, the number of agents it maintained, and the budgetary demands it placed on Whitehall were considered excessive by post-war standards. By late-1919, MI5's activities were curtailed by a reduction in staffing from over eight-hundred operatives to only twelve agents.

Although the activities of MI1(c), the Special Intelligence Service were perhaps less prominent during the Great War (due to the veil of secrecy surrounding such actions), Britain's own attempts at international espionage were a comparative success. According to the organisation itself, significant intelligence was gathered through the use of defectors and informants, and by a network of female spies known as 'La Dame Blanche' [5]. At the height of the War, La Dame Blanche was reputed to have over eight hundred operatives working on its behalf in central Europe. The group provided daily reports on the movement of German troops by rail, and were often recruited from the medical professions (as midwives and nurses had permission to cross international borders and military lines). Towards the end of the War, the SIS was awarded control of Room 40, Whitehall's signals and communications division. Room 40 was comprised of Army and Naval intelligence officers, and had deciphered a number of encrypted military and diplomatic codes during the conflict. By 1919, Room 40 was merged with Section 1b of Military Intelligence (MI1b), to form the Government Code and Cyphers School (GC&CS) operating under SIS oversight. By the early 1920s the GC&CS had managed to break the far more complicated ciphers used by the Bolshevik Government in Russia, thereby establishing a permanent legacy in British SIGINT history [4]. Yet, the most significant threat of all was entirely overlooked by the SIS during this era. Although the SIS 'uncovered evidence of Nazi-Soviet cooperation in the development of weapons technology', it was entirely unprepared for the German reoccupation of the Rhineland in 1935, preceding the outbreak of the Second World War (*Ibid*).

By the outbreak of war in 1939, MI5 and the SIS were not alone in their efforts to wage clandestine campaigns against threats to British sovereignty. During the second global conflict, the Directorate of Military Intelligence expanded both its activities and its subdivisions exponentially. Between 1939 and 1945, Military Intelligence ran nineteen such groups whose roles varied from undertaking counter-intelligence operations (MI5), debriefing prisoners of war (MI19), conducting the interrogation of escaped prisoners of war (MI9), gathering overseas intelligence (MI6/ the SIS), and directing propaganda campaigns—including the censorship of wartime correspondence (MI7). Prior to the outbreak of war, however, MI5 was considered a dysfunctional outfit despite its earlier success during WW1. This was in-part due to its adherence to out-dated operational paradigms, the sheer diversity of its portfolio of interests, due to power struggles within its leadership, and to a number of perceived failures during the Irish War of Independence. Beyond these issues, it seemed MI5 had been unable to perceive any additional threats to the UK, beyond merely hunting for spies, and according the Directorate, it could not even do that especially well.

The reason for this lapse in executive support was that during the inter-war period, Soviet intelligence services the People's Commissariat of Internal Affairs (NKVD), and the Main Intelligence Directorate (GRU), had established a new network of spies in Britain. The NKVD and the GRU gained a substantial foothold in the UK by recruiting agents directly from the nobility and from prominent public schools [3: 161–185]. MI5 was forced to adapt to these new strategies of recruitment, and upon finding out exactly who the spies were for these organisations, enlisted them to work as double agents instead. MI5's new 'XX' agents would be used to feed disinformation back to the Nazi regime in one of the War's main campaign innovations, the double cross [6]. It was the strategic use of misdirection tactics and the double cross initiative that led to the successful D-Day landings during the closing stages of the War that brought MI5 back into popular favour.

During WW2 the Government Code and Cyphers School (GC&CS) also became a key player in providing intelligence of the activities of the Nazi War Machine by intercepting important military communications. Since the First World War the GC&CS had been the UK's foremost institution for decoding diplomatic cyphers and encrypted military communications. Based at Bletchley Park near Milton Keynes, the GC&CS was responsible for breaking the German Enigma device during the Second World War, which was largely accountable for orchestrating submarine attacks against allied forces in the Atlantic Ocean. Working alongside American allies under an agreement known as UKUSA (a British/USA intelligence pact which is still in place today) the GC&CS was instrumental in the cryptanalysis of the Enigma coding device, and the later Allied victory during the Battle for the Atlantic. In 1946, the unit was renamed Government Communication Headquarters (GCHQ), and moved its centre of operations to Cheltenham, Gloucester. Here, its activities expanded exponentially in response to the emergence and wholesale popularity of electronic communication devices, and the wider deployment of Signals Intelligence (SIGINT) resources in the West. GCHQ therefore became a central competent in the UK's National Intelligence Machinery.

According to Aldrich [7: 2], two significant things happened to GCHQ towards the end of the War—mainly between 1943 and 1948. First, beyond the relocation of the GC&CS to Cheltenham and the rebranding the organisation as GCHQ after VJ-Day, a pact was made between Western Allies in the form of a new global SIGINT alliance. Consequently, this agreement was called the Five Eyes (FVEY) project as an amalgamation of SIGINT resources between the UK, the USA, Canada, New Zealand and Australia. It was named after the initial 5 stakeholders. Second, and more conspicuously perhaps, all mention of SIGINT virtually disappeared from the 'historical landscape' of intelligence programmes after the defeat of the Third Reich, and into the early days of the Cold War (*Ibid*). But this level of secrecy was not to last. In 1976, investigate journalists, Duncan Campbell and Mark Hosenball revealed the all but forgotten existence of GCHQ in an article entitled 'The Eavesdroppers' for *Time Out* magazine [8: 8–9]. Working on research that he had gathered into the UK's secret SIGINT sites (and from leaks provided by former intelligence contractors), Campbell lifted the veil on one of the UK's most secretive enterprises. Putting SIGINT into context, Campbell alleged that the UK and its intelligence allies managed a series of 'listening posts' across the world in places like 'Cyprus, Hong Kong, Singapore, Belize, Oman, St. Helena', and elsewhere (*Ibid*). In the UK, Signals Intelligence was captured via sites such as RAF Menwith Hills, RAF Irton Moor, and RAF Morwenstow in Cornwall. Not only had Campbell revealed the existence of one of the UK's most sensitive intelligence organisations, but he had discovered the sheer scale of an international SIGINT project as well. Later revelations as disclosed first by Campbell, and thereafter by former intelligence contractor and whistle-blower Edward Snowden in 2013, exposed further programmes in mass public surveillance.

This first of these projects operated under the codename of ECHELON, in which the Five Eyes group of nations (see the UKUSA pact and affiliated parties), harvested signals from the majority of telecommunications providers including military targets, business communications, and from domestic sources. Campbell initial coverage of the story was published by the *New Statesman* in [9], in which he alleged that GCHQ (now the UK's largest intelligence organisation), was intercepting all forms of telecommunications content. Of course the existence of such a project was vehemently denied by British authorities until 2001, when the European Parliament commissioned an investigation into the alleged 'existence of a global system for the interception of private and commercial communications' [10]. The European Parliament concluded that GCHQ (and the American National Security Agency, the NSA) had been intercepting domestic SIGINT in the world's largest ever surveillance project. However, it was also determined that as the main protagonists were not spying on their own countries, this particular intelligence project was not strictly breaking any laws—despite its utter contempt for personal privacy. In fact the FVEY's nations were exploiting a number of legislative loopholes in human rights doctrine that permitted them to spy on one another's populations, and then share the information gathered with the relevant parties (*Ibid*). The investigation concluded that:

> The Member States are strongly urged to review their own legislation on the operations of the intelligence services to ensure that it is consistent with the fundamental rights laid down in the ECHR and in the case law of the European Court of Human Rights and, if necessary, to adopt appropriate legal provisions. They are called upon to afford all European citizens the same legal guarantees concerning the protection of privacy and the confidentiality of correspondence. Any of their laws which are discriminatory in terms of the surveillance powers granted to the secret services must be repealed. (*Ibid*, 138)

The main findings of the commission considered that the particular types of mass surveillance being conducted had been used ‘to intercept, at the very least, private and commercial communications, and not military communications’ (*Ibid*, 11). It was established that ECHELON had been carried out (a) without the necessary permission or executive oversight required for an intelligence operation of this nature, and (b) that it was disproportionate and of questionable legality (for without identifying any specific targets, it was in violation of Article 8 of the European Convention on Human Rights). The commission called for a revision of safeguards to protect European citizens from disproportionate state surveillance, and to ensure that a necessary system of checks and safeguards be established in various EU territories to govern the security services.

However, in 2013 GCHQ became part of another international scandal, when former security contractor Edward Snowden revealed the scope and complexity of two contemporary versions of ECHELON in a monumental exposé leaked to Western news organisations. Snowden alleged that two further projects in mass surveillance had been commissioned by the FVEYs intelligence community (essentially by GCHQ and the NSA), under the operational titles of PRISM and TEMPORA [11]. It was revealed that PRISM and TEMPORA, aimed to gather intelligence from all forms of digital telecommunications streams, including those conducted online, and that data was again being processed without any legal warrant. It was alleged that GCHQ especially, had tapped into the underwater cables carrying SIGINT data between the UK and the United States. This had been done at various access points containing fibre-optic telecommunications lines throughout the UK. As much of the world’s internet traffic passes through the UK and under the Atlantic Ocean into the United States, key intelligence sites had intercepted and processed this data. Irrespective of the former ruling of the ECHELON Committee (under Article 8 of the European Convention on Human Rights), mass surveillance was being conducted on an even greater scale.

Following the disclosure of TEMPORA and PRISM by Snowden, the public outcry generated by human rights groups, by privacy advocates, and by the mass media, precipitated a number of investigations into the warrantless mass surveillance that had allegedly taken place. The main issues concerned the accumulation of Bulk Communications Data (BCD), and the second, to the interception of Bulk Personal Datasets (BPD). What this equates to is the collection of all electronic communications data by GCHQ, including the actual content of emails, of text messages, online transactions, social media profiles, internet viewing habits and chat room posts, and to the capacity of GCHQ (or its affiliates) to mine this data to profile the activities, interests, and personal relationships of anyone they deemed suspicious. In terms of the technicalities involved in an operation of this scale, it was alleged that telecommunications and internet providers had been coerced into compliance to provide access to this data, mainly due to threats restricting their legal right to trade. As a result of Snowden's revelations, rights groups Liberty and Privacy International called for a public inquiry into the alleged mass surveillance of the British population.

Following an inquest into TEMPORA headed by the UK Investigatory Powers Tribunal (set up in the wake of the ECHELON scandal during 2000), it was determined that GCHQ had acted illegally for at least seven years between 2007 and 2014, and had caused numerous human rights violations for conducting this form of mass surveillance without a warrant [12]. The general reaction from British authorities was that even if TEMPORA existed, any intercepted data would have been provided (again) by one of the FVEYs associates, thereby breaking no British laws. As the bulk of the online data in question was theoretically captured beyond Britain's territorial boundaries, this would also be acceptable. However, the Tribunal found that GCHQ had violated Sections 8 and 10 of the European Convention on Human Rights, and that further checks and balances were required to ensure the proportionality and legitimacy of the surveillance being conducted (*bid*). Following the outcome of the Tribunal, Parliament commissioned the *Investigatory Powers Act 2016*, which indeed granted these rights to British intelligence—but only on the grounds that SIGINT activities such as BCD and BPD interceptions would need to be authorised with a Home Office Warrant first. Nonetheless, the *Investigatory Powers Act 2016* retrospectively legalised the human rights violations that had been conducted by British intelligence for the previous seven years.

GCHQ Subdivisions and Capabilities

The question of what all of the above contributes to the use of modern social networks for providing security and surveillance arises from GCHQ's most recent SIGINT innovations under programmes such as TEMPORA. In 2003, the Joint Threat Analysis Centre (JTAC) was established to work alongside MI5 in the pursuit of terrorists and in the prevention of terrorism-related attacks and serious organised crime. The JTAC is provided with intelligence gathered by organisations such as GCHQ, and disseminates material of interest to other security providers including MI5, MI6, and Counter-Terrorism Command at the Metropolitan Police Service. Also within the UK's National Intelligence portfolio, is a little-known organisation working under the management of GCHQ. The Joint Threat Research Intelligence Group (JTRIG) was revealed as part of the disclosures published in 2013 by journalist Glenn Greenwald and by Edward Snowden. Snowden's documents revealed that the JTRIG operates as part of GCHQ's clandestine SIGINT programme and provides a contemporary rendition of MI5's counter-subversion activities as seen during the two global conflicts of the last century.

Essentially, the JTRIG exploits Signals Intelligence provided by GCHQ, though it is also capable of detecting its own targets and for managing its own portfolio of interests. As opposed to working with Covert Human Intelligence Sources (C.H.I.S.), as might be the case for MI5 and MI6, the JTRIG focuses exclusively 'on the cyber domain (computers and the internet), using both open source data and SIGINT' [13]. In terms of the basic dimension of these activities, the JTRIG has been reputed to conduct SIGINT, COMINT, and COINTELPRO operations, and has ventured into more clandestine territory in terms of directing Human Intelligence (HUMINT) and conducting Psychological Operations (PSY-OPS). The purpose of these techniques according to the JTRIG is to 'make something happen in the real or cyber world', or in other words, to produce Operational EFFECTS [14]. The JTRIG's Operational EFFECTS programmes can be considered in a number of ways, but principally these include (a) the use of persuasive tactics to change a suspect's potential behaviour; (b) the infiltration of public, private, and secure networks online to monitor the activities of designated groups; (c) where necessary to intervene or plant material to degrade an individual's reputation; and (d), to misinform, deceive, or distract their targets using online propaganda. Indeed all of the above pays homage to the origins of British military intelligence. The JTRIG achieves all of the above, through its exploitation of social networks using platforms such as YouTube, Facebook, private email and chat room clients, and through its use of False Flag events or news items released via the internet into the public domain.

During 2011, prominent psychologist Dr. Mandeep K. Dhami was seconded to the Human Systems Group, Information Management Department at the Defence Science and Technologies Laboratory (DSTL) at Porton Down. Porton Down is one of the British Army's most sensitive establishments and is used as a research facility into chemical, biological, radiological, and nuclear (CBRN) defence initiatives, and for the associated military application of a range of other issues as well. In March 2011, Dhami published a classified document entitled *Behavioural Science Support for JTRIG's (Joint Threat Research and Intelligence Group's) Effects and Online HUMINT Operations*, which documents many of the JTRIGs operational capabilities. Ostensibly, the JTRIG's activities include the use of online techniques to:

> discredit, disrupt, delay, deny, degrade, and deter

The actions of suspects identified by SIGINT operations [15]. How the JTRIG conducts its military and law enforcement actions can be considered as follows, by:

- Uploading YouTube videos containing "persuasive" communications (to discredit, promote distrust, dissuade, deter, delay or disrupt)
- Setting up Facebook groups, forums, blogs and Twitter accounts that encourage and monitor discussion on a topic (to discredit, promote distrust, dissuade, deter, delay or disrupt)
- Establishing online aliases/personalities who support the communications or messages in YouTube videos, Facebook groups, forums, blogs etc.

- Establishing online aliases/personalities who support other aliases
- Sending spoof e-mails and text messages from a fake person or mimicking a real person (to discredit, promote distrust, dissuade, deceive, deter, delay or disrupt)
- Providing spoof online resources such as magazines and books that provide inaccurate information (to disrupt, delay, deceive, discredit, promote distrust, dissuade, deter or denigrate/degrade)
- Providing online access to uncensored material (to disrupt)
- Sending instant messages to specific individuals giving them instructions for accessing uncensored websites
- Setting up spoof trade sites (or sellers) that may take a customer's money and/or send customers degraded or spoof products (to deny, disrupt, degrade/denigrate, delay, deceive, discredit, dissuade or deter)
- Interrupting (i.e., filtering, deleting, creating or modifying) communications between real customers and traders (to deny, disrupt, delay, deceive, dissuade or deter)
- Taking over control of online websites (to deny, disrupt, discredit or delay)
- Denial of telephone and computer service (to deny, delay or disrupt)
- Hosting targets' online communications/websites for collecting SIGINT (to disrupt, delay, deter or deny)
- Contacting host websites asking them to remove material (to deny, disrupt, delay, dissuade or deter) (*Ibid*: 9)

In addition to Dhami's account of the JTRIG's activities, leaked intelligence documents from *CBS News*, and *The Intercept* (largely gathered from the Snowden archives), have detailed a number of case studies in which JTRIG EFFECTS Operations have been a success. Of the JTRIG's main subdivisions (the Iran Team, the Serious [Cyber] Crime Team, the Global Team, the Counter-Terrorism Team, the Cyber Coordination and Operations Team, and the Network Defence Team), the organisation is able to cover the majority of contingencies that might arise online.

For example during 2011, the Global Team was "monitoring" both the impending regime change in Zimbabwe 'by discrediting the present regime', and it was also 'preventing Argentina from taking over the Falkland Islands by conducting online HUMINT' (*Ibid*: 8). The JTRIG's latter operation included posting online content to dissuade Argentinian voters from supporting their government's intentions to reclaim the Falkland Islands under Argentine sovereignty. The Serious [Cyber] Crime Team was using EFFECTS protocols to reduce consumer's trust in both 'front companies' and those selling counterfeit online goods (*Ibid*). The Cyber Coordination and Operations team, was comparatively investigating 'cybercrime and electronic attack[s] by: (1) denying, deterring or dissuading criminals, state actors and hacktivists; (2) providing intelligence for judicial outcomes; and (3) discrediting cybercrime forums and their users' (*Ibid*). However, in relation to potential infringements on freedom of speech (especially those cited under article 10 of the ECHR—the right to freedom of speech), in 2011 the Serious [Cyber] Crime Team was involved with monitoring 'domestic extremist groups' and other potential threats to Parliamentary Democracy and public safety in the UK (*Ibid*: 9).

The following section of this chapter, examines how some of these techniques have been used to prevent serious organised crime, and as a means to maintain public order.

JTRIG COINTELPROs

To briefly put the JTRIG's Operational EFFECTS into some form of context, a directory of tools used by this organisation was published online via *The Intercept* [16]. The dossier included screen shots of the catalogue of tools that JTRIG agents are able to deploy in any given scenario [17]. These include EFFECTS programmes such as Angry Pirate, Rolling Thunder, and Vipers Tongue, and a series of tools used in the forensic investigation of remotely accessed computers. Angry Pirate for example, is reputed to 'disable a targets account on a computer', whereas Rolling Thunder can be used to conduct a distributed denial of service (DDoS) assault, which is generally deployed to take a website offline [16]. Comparatively, the Vipers Tongue EFFECTS programme is used to prevent calls being made or received by satellite and GSM mobile phones. Packages such as SNOOPY and BEARSCAPE are respectively used to replicate data sets from mobile phones and to plunder the handsets for Wi-Fi connection history.

In relation to known instances in which EFFECTS programmes have been used by the JTRIG, Hacktivist groups such as Anonymous have proven to be a viable target against which to test some of these systems. During Operation Payback in 2011 for example, the JTRIG used HUMINT techniques to coerce one such Hacktivist into divulging his haul of sensitive stolen data from the Federal Bureau of Investigations [18]. Here, agents (possibly from JTRIG's Cyber Coordination team) infiltrated an IRC known to be used by Hacktivists, and waited for an individual to start boasting about his antics. When a hacker named p0ke announced that he had stolen the details of over 700 employees from a government website, the agent asked if he had seen a BBC website entitled 'Who loves the hacktivists'. The website was a snare designed to track internet traffic to the page, and this was used to trace p0ke's IP address—leading to a conviction.

> When p0ke clicked on the link […] JTRIG was able to pull up the IP address of the VPN (virtual private network) the hacktivist was using. The VPN was supposed to protect his identity, but GCHQ either hacked into the network, asked the VPN for the hacker's personal information, or asked law enforcement in the host nation to request the information (*Ibid*).

In relation to JTRIG Operational EFFECTS, the purpose of this exercise was to assist law enforcement officials to make an arrest of the suspect, who by his own admission had provided all of the evidence required to make a conviction.

During another 2011 campaign (codenamed Operation Wealth), the JTRIG experimented with a new technical protocol entitled Rolling Thunder. According to the JTRIG, the purpose of Rolling Thunder was to destabilise communications between groups of Hacktivists belonging to the Anonymous collective. In this operation, the JTRIG targeted an Internet Relay Chat room (IRC) which was being used by ‘Anonymous, LulzSec and the Syrian Electronic Army’ [19]. Here, Rolling Thunder was used to launch a coordinated DDoS assault against the IRC, as a means to prevent Hacktivists from organising further attacks on corporate or public sector networks. Ironically, the techniques used by Hacktivists to disable commercial and state websites had been turned against them. Primarily Operation Wealth was intended to provide intelligence to law enforcement agencies of who the key protagonists were in the movement. However, the operation also sought to disrupt one of the group’s main communications platforms, thus to highlight the vulnerability of this particular medium to Anonymous. According to journalist Gerry Bello:

> The JTRIG infiltrated chat rooms and other online social spaces used by Anonymous to gain human intelligence on Anonymous members. Once gaining the hacktivist’s trust they inserted spyware called Spyeye. This spyware replicated across many computers, converting them into a single remote controllable network entity [20].

The above network of bots (remotely accessed computers) was thereafter used to implement the DDoS assault, rendering the IRC utterly useless. In terms of operational EFFECTS, 80% of the users engaged in this forum had abandoned the site 1 month later according to the JTRIG [18]. Nevertheless, according to Bello, the site was also being used by legitimate social movements as well, in terms of providing a secure and (presumably) private space in which to organise public campaigns. Taking the IRC offline potentially caused an infringement of the Human Right to Freedom of Speech for such groups.

Further examples of the particular types of activity that can be accredited to the JTRIG include a number of potential COINTEL and HUMINT operations (and their EFFECTS) that were conceivably conducted during the Million Mask March in 2015. The Million Mask March has rapidly become an annual event since it first started in 2013; in which campaigners representing a wide variety of social issues converge on various state capitals to demonstrate against state surveillance, government and corporate corruption, austerity measures, financial inequality, and the securitization of the internet. The movement has gained a substantial following, with international demonstrators marching through the streets of major towns and cities every November 5th wearing the now-familiar Guy Fawkes “Guido” masks, synonymous with the Anonymous collective and the dystopian epic, *V for Vendetta*. During the 2015 Million Mask March in London, a number of conspicuous events took place. First the organisers of the event drew attention to a fake Facebook page and to a website that had emerged directing participants to alternative locations for the protest. Then, the website for London’s Metropolitan Police Service was taken offline, arguably due a DDoS attack committed by Anonymous. At one point during the protest a police car was also set alight allegedly by demonstrators, though almost immediately the internet was awash with claims that the vehicle was a hoax, or worse still, a plan by the authorities to depict the event as a riot [13].

The point is that sometimes it is just enough to provide a distraction from the real events to gain a small margin of success. With regards to online speculation about the "false flag" burning of a police car, it was suggested that the car itself was a propaganda stunt. For several days after the demonstration, the denizens of the internet spent hours looking for evidence that the car was broken-down wreck that had been towed to the site, but of course, promoting these kind of stories keeps people preoccupied and prevents them from getting up to mischief elsewhere online [21]. Indeed those visiting or commenting on such pages can also be tracked in relation to their particular take on events.

Ethical and Legal Considerations

A far bigger question then, relates to the ethical and legal issues posed (a) by the net cast by these sweeping forms of mass surveillance, and (b) the operational protocols themselves, in so far as they aim (in some cases) to manipulate public opinion on heated political debates, and may change the outcome of policymaking decisions without due consent. With regards to the legislative factors that enable mass surveillance, there are two issues of concern. The first of these issues regards the public exposure of such programmes, is as much as information regarding military and domestic surveillance operations is highly classified, and is protected as a matter of National Security. The second factor to be considered is the legitimacy of such operations in relation to Human and Civil rights affairs. Of course both of the above issues are intrinsically bound in terms of public policy, under which a number of contingencies are designed to maintain official secrecy. Nevertheless, the question of the legitimacy of state surveillance depends entirely on the population's knowledge of it.

According to The *Investigatory Powers Act 2016*, communications interception warrants (including all forms of digital and telecommunications data transmitted in the UK), can now only be requested by the most senior officials in the security services, in military intelligence, or by the police. These measures are intended to represent a safeguard against the unrestrained and unwarranted use of mass surveillance. However, under the *Investigatory Powers Act 2016*, the number of instances in which social movements can be identified as posing a potential treat, i.e.; to public safety, for preventing outbreaks of disorder, and even for defaming the economic interests of the United Kingdom, indicates that a range of campaign organisations are the potential targets for surveillance. Under the *Act*, the power to grant authorisation for obtaining communications data will usually be given by the Secretary of State providing they meet the following criteria:

(a) in the interests of national security,
(b) for the purpose of preventing or detecting crime or of preventing disorder,

(c) in the interests of the economic well-being of the United Kingdom so far as those interests are also relevant to the interests of national security,
(d) in the interests of public safety,
(e) for the purpose of protecting public health,
(f) for the purpose of assessing or collecting any tax, duty, levy or other imposition, contribution or charge payable to a government department,
(g) for the purpose of preventing death or injury or any damage to a person's physical or mental health, or of mitigating any injury or damage to a person's physical or mental health,
(h) to assist investigations into alleged miscarriages of justice,
(i) where a person ("P") has died or is unable to identify themselves because of a physical or mental condition

    (i) to assist in identifying P, or
    (ii) to obtain information about P's next of kin or other persons connected with P or about the reason for P's death or condition, or

(j) for the purpose of exercising functions relating to

    (i) the regulation of financial services and markets, or
    (ii) financial stability [22].

Overall, the ethical considerations posed by the *Investigatory Powers Act 2016* would seem to be relatively straightforward, in as much as certain types of surveillance may only be conducted with prior authorisation, and for very particular reasons. However, in terms Human and Civil Rights concerns, what the question really becomes is how are these issues represented or defined in the Act?

In this respect, Section 231 of the *Investigatory Powers Act 2016* focuses on the ethical issue of 'Error reporting', that is, if an operative believes a human right has been violated—then how should this be voiced? The Act states that:

> The fact that there has been a breach of a person's Convention rights (within the meaning of the Human Rights Act 1998) is not sufficient by itself for an error to be a serious error (*Ibid*: 187).

Intelligence officers are, therefore, only required to disclose human rights violations (as might be in the public interest for example), providing they do not lead to a lack of confidence in 'national security', to disrupt the 'the prevention or detection of serious crime', prove detrimental to 'the economic well-being of the United Kingdom', or impede 'the continued discharge of the functions of any of the intelligence services' (*Ibid*). Accordingly, this contingency presumes that human rights violations will take place in the future, and prevents the general public from finding out. An addition concern to be cited then, relates to how social movements in the UK, are now framed in National Security policy and in criminal law, to become aligned with the above high risk categories of national security, serious organised crime, public health, and public safety.

Re-defining Terrorism, Subversion, and Dissent

In today's SIGINT market there is perhaps less of an emphasis being placed on Military Intelligence (though make no mistake such institutions collate data from all major communications streams), and a revised focus on detecting new targets for security agencies to track. The main shift from gathering Military Intelligence, to conducting domestic surveillance (or rather, for providing real-time situational awareness of important communications data and events), relates to how the role of institutions such as MI5 and GCHQ has evolved over the last hundred years. Arguably many of these changes took place in secret during the closing stages of the Cold War amidst the collapse of the Soviet Bloc, with a renewed interest being shown in both political subversion and in domestic affairs such as the maintenance of public order. According to Andrew [3], the expansion of Military Intelligence agencies into conducting domestic surveillance, originally formed part of MI5's attempts to legitimise the growth of the organisation during the Great War. Comparatively, Stella Rimmington (the former Director-General of MI5), has indicated that MI5 has historically maintained an interest in groups who have aimed to subvert British Parliamentary Democracy, demonstrating that the particular types of surveillance used for military purposes have a concise domestic application as well (2001: 161).

During the 1980s especially, MI5 was alleged to have infiltrated the Campaign for Nuclear Disarmament (CND), and was suspected of running a series of smear campaigns in the popular press to discredit the leadership of the National Union of Miner's [23: 307–308]. MI5 suspected that the communist party was intending to infiltrate the CND and use it as a front to destabilise the constitutionally elected Government of Margaret Thatcher, and thus posed a threat to National Security. Rimmington has also claimed that the triumvirate of leaders behind the National Union of Miner's (NUM) had openly expressed an intention to destabilise the government, thereby categorising the movement as an immediate target for state surveillance (2001: 163). However, this particular counter-subversion role of MI5 is rather more indicative of its role during peacetime, in which gathering Military Intelligence is substituted for maintaining National Security.

Today, MI5 is responsible to the Home Office, whereas MI6 is responsible to the Foreign Office. Little has changed in this particular context. Nonetheless, following the Cold War, the role of MI5 was somewhat refined, and again this reflects the way in which threats to the UK are now portrayed. In fact the existence of both MI5 and MI6 was only formally recognised by Whitehall in the *Secret Services Act 1989*. Comparatively, law enforcement activities were only included within MI5's operational remits as of 1996. In relation to supporting the police in the pursuit of serious organised crimes, MI5 currently defines its role as being the promotion of National Security against attacks from terrorist organisations, against industrial sabotage and espionage, protecting the UK against foreign states or agents, and 'from actions intended to overthrow parliamentary democracy' (MI5 2017). It is the above threats to National Security, through which social movements in the UK have been reclassified as targets for state surveillance, the legitimising factors of which can be found in British law.

Post-9/11 public policy has identified that groups intending to conduct campaigns at designated sites of Critical National importance to the UK (such as power stations and utilities providers, transportation networks and the financial infrastructure), are considered threats to National Security under the UK's Civil Contingencies Programme. Under the *Civil Contingencies Act 2004*, it became an offence to interfere with the commercial operations of such sites, and a range of civil contingencies partnerships were set up to identify key risks to the Critical National Infrastructure [24]. The *Civil Contingencies Act* was originally intended to replace the out-dated *Defence Act 1948* [25], and the *Emergency Powers Act 1920* [26]. It was considered that existing legislation had failed to cope with events such as the outbreak of foot and mouth disease in 2001 and the nation-wide fuel protests of 2000—both of which had threatened the provision of crucial domestic services in the UK. The Act was also designed to defend the Critical National Infrastructure of the UK from the threat posed by international terrorist groups committing violent attacks on British soil.

A similar shift occurs in public order discourse at this particular point in history, in which the terminology used to define both terrorists and activists substantially changed. Around 2004, a number of think tanks in the UK and USA started to adopt a new set of descriptors, through which to define terror attacks and those associated with the activities of radical social movements. It was here that terrorists were redefined as violent extremists, and political activists as domestic extremists [27]. In National Security discourse it was considered that using terms such as 'jihad' legitimised a belief in anti-western values, but again the relationship between defence and security demonstrates how the intelligence community justifies its surveillance of social movements in the UK. As noted by Monaghan and Walby [28], Jones [29], and Harbisher [27], the notion of domestic extremism has gained popularity across the FVEYs intelligence community, and it is used to defame any groups that promote anti-Western sentiments, or who vociferously refute the decisions of Western policy makers. In Monaghan and Walby's work, the Canadian Integrated Threat Assessment Centre (Canada's equivalent of the UK's JTAC), defined campaign groups in opposition to the 2010 Vancouver Olympic Games as an extremist threat [28: 148]. Torin Monahan has also observed that similar threat assessment centres in the United States of America have subjected race orientated groups and student associations to unprecedented levels of surveillance—as they are considered potential hotbeds for extremism [30: 48].

In the UK, international campaign groups such as the Occupy movement have equally been depicted by the security services as domestic extremists. This technique forms a particular discourse which serves to legitimise how the modern state seeks to control public demonstrations. For example, during the 2011 occupation of the steps outside St. Paul's Cathedral, campaigners became aware that their mobile devices had stopped working [1]. Legal representatives for the group later filed a request under the *Freedom of Information Act 2000*, in an attempt to determine whether or not GCHQ's anti-terrorism capabilities had been used against them to end the demonstration before the 2012 London Olympic Games [31]. Of course it is not unusual for Britain's security services to undertake collaborative ventures in this manner. During her work for the DTSL, Mandeep Dhami also alludes to these collaborative partnerships:

> Within GCHQ, the teams work with the relevant Intelligence Production Teams (IPTs) who aid in the initiation and planning of operations based on their analysis of SIGINT … Several teams currently collaborate with other agencies including the SIS, MoD's Technical Information Operations (TIO), the FCO, Security Service, SOCA, UK Borders, HMRC, Metropolitan police, and the National Public Order and Intelligence Unit [15: 6].

So not merely is the securitization of cyberspace a question of how modern security agencies define the threats they perceive to British democracy, but in relation to the Operational EFFECTS enacted by groups such as the JTRIG, responses to such issues are conducted using increasingly militaristic techniques that pay homage to the wartime origin of these organisations. The most chilling impact of all is the prospect that on an increasing scale, members of various social movements are becoming aware of what they might say or do online, and how this might be used against them. Recent news reports have indicated that citizens of the democratic West have started to censor their own internet posts for fear of what might happen to them if they openly voice their concerns [32].

Conclusion—Risks and Recommendations

To elaborate on the above problematic, the widespread coverage of state surveillance programmes such as PRISM and TEMPORA in the media, and the documented use of COINTEL procedures against campaigners, has had a number of chilling effects. The first of these (as observed above) relates to the reclassification of social movements (such as Occupy and various Environmental groups) as domestic extremists in National Security policy. The second concern is the legitimacy this provides to intelligence agencies to conduct communications surveillance over any groups or individuals they perceive as posing a threat—seemingly with limited consequences for violating their human rights by doing so. Yet, the biggest concern of all is the scant differentiation between legitimate threats and expressions of opinion, which are now determined according to risks posed to the Critical National Infrastructure, or to parliamentary democracy, or in other words—to threats which are not of a terrorist nature. This latter issue frames the conclusion for this Chapter, in as much as for campaigners, there is now significant awareness within these communities of the scope of this surveillance, which is causing them to be cautious in expressing their opinions online.

Recent academic studies have, therefore, attempted to demonstrate precisely how online surveillance has started to shape dissenting public discourse. Stoychef, for example, has typified this concern in an article entitled 'Under Surveillance:

Examining Facebook's Spiral of Silence Effects in the Wake of NSA Internet Monitoring' by suggesting that:

> In today's Internet age, the expression of online opinions leaves digital footprints, inextricably linking individuals to political views they shared weeks, months, and even years prior. In other words, there is a newfound permanency associated with a one-time willingness to speak out online [33: 289].

The study provides an analysis of the readiness of the American public to debate key human and civil rights issues on social media, specifically those pertaining to surveillance programmes such as PRISM. The review concluded that whereas '86% of respondents were willing to discuss the Snowden PRISM leak in offline settings […] less than half of those would post about it on Facebook or Twitter' (*Ibid*: 299). Indeed for 'the remainder—and majority—of participants, being primed of government surveillance significantly reduced the likelihood of speaking out in hostile opinion climates' such as those found online (*Ibid*).

In other reported cases in which cyber-censorship has had a detrimental impact on freedom of speech, various news agencies throughout the world have concluded that COINTEL Operations are being used in a similar manner. In Belarus for example, state surveillance technology has removed open access to 'particular websites including Facebook' and has enabled 'the creation of fake versions of popular dissident websites' [34]. In South East Asia, Shetty has observed that Vietnamese activists have 'retreated to the internet' only to discover that 'with invasive surveillance regimes in place and a proliferation of new laws that govern online offences, there are few places left for people to gather, speak or write as they wish' [35]. In the UK, a recent study has also determined that activists are equally aware of the different types of online surveillance in which they participate. According to Ramsay et al. knowledge of state 'surveillance is largely conditioned by the specific experiences of groups and individuals and by the immediate need to perform successful acts of protest' [36]. In their recent study of public order interventions in the UK, campaigners expressed a greater concern at overt forms of surveillance (i.e.; the use of police cameras and personal searches), and the potential for physical coercion to be enacted during demonstrations, than they did about online surveillance. However, when asked specifically about intelligence organisations such as GCHQ and the JTRIG:

> News of this programme appeared to resonate with activists' direct experiences, such as the sudden appearance of disruptive trolls in online discussions (2017).

Nonetheless, one of the main observations made by the above was the relatively low levels of security employed by campaigners during public demonstrations in the UK. In relation to the alleged use of COINTEL practices during the Occupy London protests (to disrupt the communications devices of campaigners), there is now a demonstrable need for enhanced levels of security to be used during public campaigns [1]. According to Lee and Feeney:

> Police often spy on protesters, and the smartphones they carry, and no matter how peaceful the demonstration, there's always a chance that you could get detained or arrested, and your devices could get searched [37].

Therefore, in a society in which smartphones can be remotely hacked, in which the digital footprint of one's activities remain online, in which secure digital networks get accessed by the state, and in which fundamental human rights seem increasingly at stake, private citizens and campaigners alike should take their privacy far more seriously. This proposition seems increasingly led by privacy groups and by members of the free press, who in the last couple of years have started to promote encryption methods that may well become the norm. Lee and Feeney especially, are now part of this debate in as much as they offer 'tips on how to prepare your phone before you go to a protest' (*Ibid*). Generally speaking this should be seen as good advice.

References

1. Occupy London (2015) City police questioned on likely use of GCHQ's "anti-terror" capabilities during policing of Occupy London. http://occupylondon.org.uk/domestic-extremism/. Accessed 23 Nov 2015
2. Cabinet Office (2008) The national security strategy of the United Kingdom. Cabinet Office, London
3. Andrew C (2012) The defence of the realm. Penguin, London
4. Lerner (2017) MI6 (British secret intelligence service). Espionage Information. http://www.faqs.org/espionage/Lo-Mo/MI6-British-Secret-Intelligence-Service.html. Accessed 02 Jan 2017
5. SIS (2017) MI6: our history. https://www.sis.gov.uk/our-history.html. Accessed 02 Jan 2017
6. The National Archives (2017) Double cross agents. http://www.nationalarchives.gov.uk/releases/2003/may22/doublecross.htm. Accessed 02 Jan 2017
7. Aldrich R (2010) GCHQ: the uncensored story of Britain's most secret intelligence agency. Harper Press, London
8. Campbell D, Hosenball M (1976) The eavesdroppers. Time Out, pp 8–9
9. Campbell D (1988) Somebody's listening. New Statesman
10. Schmid G (2001) Report on the existence of a global system for the interception of private and commercial communications (ECHELON interception system). European Parliament, Brussels
11. Shubber K (2013) A simple guide to the prism controversy. Wired. http://www.wired.co.uk/article/simple-guide-to-prism. Accessed 05 Sept 2010
12. Bowcott O (2015) UK–US surveillance regime was unlawful 'for seven years'. The Guardian. https://www.theguardian.com/uk-news/2015/feb/06/gchq-mass-internet-surveillance-unlawful-court-nsa. Accessed 23 Nov 2015
13. Harbisher B (2016) The million mask march: language, legitimacy, and dissent. Crit Discourse Stud 13(3): Discourse and Protest Events
14. Greenwald G (2014a) How covert agents infiltrate the internet to manipulate, deceive, and destroy reputations. The Intercept. https://theintercept.com/2014/02/24/jtrig-manipulation/. Accessed 02 Jan 2017
15. Dhami K (2011) Behavioural science support for JTRIG's (Joint Threat Research and Intelligence Group's) effects and online HUMINT operations. DTSL, Porton Down

16. The Intercept (2014) JTRIG tools and techniques. https://theintercept.com/document/2014/07/14/jtrig-tools-techniques/Cited Accessed 02 January 2017
17. Greenwald G (2014) JTRIG tools and techniques. The Intercept. https://theintercept.com/document/2014/07/14/jtrig-tools-techniques/. Accessed 02 Jan 2017
18. Schone M (2014) Exclusive: Snowden docs show UK spies attacked anonymous, hackers. NBC. http://www.nbcnews.com/news/investigations/war-anonymousbritish-spies-attacked-hackers-snowden-docs-show-n21361. Accessed 02 Jan 2017
19. Gilbert D (2014) UK government used 'rolling thunder' DDoS attacks against anonymous, LulzSec and Syrian electronic army. International Business Times. http://www.ibtimes.co.uk/uk-government-used-rolling-thunder-ddos-attacks-against-anonymouslulzsec-syrian-electronic-1435186. Accessed 02 Jan 2017
20. Bello G (2014) Anonymous revealed to be under attack from both corporate and government spies. Free Press. http://freepress.org/departments/display/20/2014/5336. Accessed 02 Jan 2017
21. West M (2015) Debunked: cop car towed to media location then torched at million mask march—BX10 LNV. Metabunk. https://www.metabunk.org/debunked-copcar-towed-to-media-location-then-torched-at-million-mask-March-bx10-lnv.t6962/. Accessed 02 Jan 2017
22. HMSO (2016) Investigatory powers act 2016. Her Majesty's Stationary Office, London
23. Milne S (2004) The Enemy within. Verso Books, London
24. HMSO (2004) Civil contingencies act 2004. Her Majesty's Stationary Office, London
25. HMSO (1948) Defence act 1948. Her Majesty's Stationary Office, London
26. HMSO (1920) Emergency powers act 1920. Her Majesty's Stationary Office, London
27. Harbisher B (2015) Unthinking extremism: radicalising narratives that legitimise surveillance. Surveill Soc 13(3/4):474–486
28. Monaghan J, Walby K (2012) Making up 'Terror Identities': security intelligence, Canada's integrated threat assessment centre and social movement suppression. Polic Soc: Intl J Res Policy 22(2):133–151
29. Jones C (2014) "Call it intercontinental collaboration": radicalisation, violent extremism and fusion centres. Statewatch, London
30. Monahan T (2010) The future of security? Surveillance operations at homeland security fusion centres. Soc Just 37(2–3):84–98
31. RT News (2015) Police told to apologize for treating Occupy London protesters as extremists. RT News. https://www.rt.com/uk/314846-occupy-london-protest-surveillance/. Accessed 02 Jan 2017
32. Turner K (2016) Mass surveillance silences minority opinions, according to study. Washington Post. https://www.washingtonpost.com/news/the-switch/wp/2016/03/28/mass-surveillance-silences-minority-opinions-according-to-study/. Accessed 02 Jan 2017
33. Stoychef E (2016) Under surveillance: examining facebook's spiral of silence effects in the wake of NSA internet monitoring. J Mass Commun Q 93(2):296–311
34. Taylor J (2013) Government of belarus using 'new tools' to silence dissent on internet, says index on censorship report. Independent. http://www.independent.co.uk/news/world/europe/government-of-belarus-using-new-tools-to-silence-dissent-oninternet-says-index-on-censorship-report-8438880.html. Accessed 23 Nov 2015
35. Shetty S (2016) Why is East Asia silencing free speech. World Econ Forum. https://medium.com/world-economic-forum/why-is-east-asia-silencing-free-speech-663726d1b0c8. Accessed 02 Jan 2017
36. Ramsay G, Ramsay A, Marsden S (2016) Impacts of surveillance-on-contemporary British activism. Opendemocracy. https://www.opendemocracy.net/uk/gilbert-ramsay/report-impacts-of-surveillance-on-contemporary-british-activism. Accessed 23 Nov 2015
37. Lee M, Feeney L (2017) Cybersecurity for the People—how to protect-your privacy at a protest. Intercept. https://theintercept.com/2017/04/21/cybersecurity-for-the-people-how-to-protect-your-privacy-at-a-protest/. Accessed 23 Nov 2015

Chapter 14
Surveillance, Targeted Killing and the Fight Against Terrorism: A Fair-Played Game?

Ioanna K. Lekea

Introduction

After the attacks of 9/11 and the subsequent war against terrorism, many questions arose concerning the difficulties of observing international humanitarian law during asymmetrical warfare. On one hand, terrorists do not pay attention to the Geneva Conventions or any other treaties concerning the respect of human rights, the protection of non-combatants or the permissible means of fighting; instead, they attack innocent people to accomplish their goals and put pressure on their opponents.

On the other hand, just war theory (hereafter referred to as JWT), along with other arguments based on the violation of human rights, seem to have played an important role on the decision to declare the 'war against terror'. Initially, the justification of the war was based on the self-defence argument [1, 2]. In the course of time, another reason that was used as a justification was that the Taliban regime was suppressing fundamental human rights [3], such as the peoples' right to (religious) freedom, to a just trial, to freedom of speech and act—let alone women's rights to education and self-determination.

In theory, it seems as if the rules and the obligations of the military are well-defined and expressed in the existing Rules of Engagement (hereafter referred to as RoE) and International Humanitarian Law (hereafter referred to as IHL); members of the armed forces are obliged to observe the law and protect non-combatants, treat the prisoners of war well, and comply with the law, even if the enemy does not. However, in practice things are not quite so simple both on legal and ethical grounds [4, 5].

I.K. Lekea (✉)
Division of Leadership-Command, Humanities and Physiology, Dekeleia Air Base,
Department of Aeronautical Sciences, Hellenic Air Force Academy, Attiki 13671 (1010),
Athens, Greece
e-mail: ioannalekea@gmail.com; ioanna.lekea@hafa.haf.gr

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_14

Scope of This Chapter

The main purpose of this chapter is to illustrate the difficult ethical questions [6] raised by the conduct of asymmetrical war. IHL, as well as the principles of JWT, can provide us with the tools necessary not only to evaluate ethical questions on the use of force, but also to prepare us to deal with practical issues about the conduct of hostilities [7]. The issues we will be addressing in this chapter can be summarized in the following questions:

- If the enemy is not acting by the law [8, 9], could one possibly support the argument that lawful combatants are also justified to do the same, thus vindicating and legitimizing illegal and immoral behaviour, which is not in conformity with IHL and the UN Charter? We will be using the targeted killing strategy as a case study to answer this question, as it is a much debated issue and a good tool to look at the extent to which the war against terrorism is conducted on fair grounds. The policy of targeted killing is of special interest as it seems to be a very popular anti-terror policy and has implications at different levels [10]: IHL, human rights and the ethical conduct of hostilities [11]. In the following sections we will be examining both the legality and the morality of this practice in relation to issues of (non) privacy and surveillance [12, 13] of individuals and groups of peoples based on their beliefs, religion, nationality or country of residence.
- When the armed forces decide to set aside the obligation to observe fundamental human rights [14] using surveillance against suspects and other techniques that might prevent a terrorist attack from taking place [15, 16], but at the same time violate peoples' privacy and freedom to communicate or act, what is the benefit and the overall damage [17]?

Targeted Killing: The Moral and Legal Framework

The strategy of killing certain individuals when they are not posing any direct threat [10] (targeted killing [18]) resembles the policy that Israel [10] applies against Palestinians and seems to be adopted by the US Armed Forces. The American political and military authorities opposed that policy at the beginning—however, they actually follow it by pre-emptively killing terrorist suspects [19]; before deciding whether they should kill them or not, they deprive them of their privacy since being a suspect means that the identified individual's life is under surveillance [20]. For the purposes of this study, the definition of targeted killing in the context of the war against terrorism can be given as the decision to kill a suspect for planning and intending to execute terrorist acts, if he/she presents a serious threat to public order (according to criminal evidence and intelligence information) with the ultimate aim to cancel, prevent or at least limit future terrorist attacks [20].

This should be the last resort, where all other methods of incapacitating him/her have been tested and failed [21].

Many cases could be used as paradigms, such as the Israeli attack on Sheikh Yassin of Hamas or the targeted killing of a British citizen in Syria by a RAF drone strike on 21 August 2015. The cases discussed in this chapter are typical (and thorough examined) examples of how targeted killing has been used, as a tactical option, in the early stages of the war against terrorism and onwards [22–26].

In this context, we chose to refer to two characteristic operations of that kind: (1) the attack against six terrorist suspects in Yemen on November 2002 [27, 28]; and (2) the attack that took place in June 2006 and lead to the death of Abu Musab al-Zarqawi in Iraq. In the first case, a US missile attack launched from a Predator drone aircraft targeted six suspected Al Qaeda terrorists travelling in a vehicle to Yemen. Although civilians were killed, the attack was considered a military action and not an assassination [29]. As a result of this attack Al-Harithi, who was the leader of many Al Qaeda attacks, was killed.

The attack against Abu Musab al-Zarqawi was justified on similar grounds. Zarqawi was thought to be the mastermind of numerous acts of violence in Iraq, including suicide bombings and hostage executions. He was believed to be responsible for co-ordinating numerous suicide bombers throughout Iraq. Zarqawi was killed on June 7, 2006, while attending a meeting in an isolated safehouse approximately 8 km north of Baqubah. Two United States Air Force F-16C jets identified the house and bombed the building. Except for Zarqawi, six other individuals were also reported killed. The death of one of his key lieutenants, spiritual adviser Sheik Abd-Al-Rahman was announced at a later date.

In general terms, three conceptual frameworks can be used to decide on the legitimacy and morality of targeted killing of terrorist suspects at the level of conducting counter-insurgency warfare:

- To treat it as a legal and morally acceptable operation [27];
- To treat it as an act that violates fundamental human rights and that can be justified only on the grounds of very strict criteria and in special circumstances [18];
- To treat it as an act that is morally unacceptable, as any kind of violence, and should be condemned even if carried out in the course of a war. This is the position of those who adhere to pacifism and I mention it here only for reasons of completeness. As it is not realistic to expect that there would ever be a war without casualties or that in the near future military operations would cease to exist, this particular position will be no further analysed.

According to the first opinion, terrorists should be dealt with as lawful combatants, who, even if they do not wear uniforms or they do not bear any insignia, retain their status for as long as they are directly involved in military operations. Their classification as lawful combatants is, however, problematic as IHL states that protection should be given to those who: have a person in charge responsible for the acts of the people under his command, have a distinctive badge all the time which can be recognised from far away, carry weapons overtly and comply to the laws and the customs of war in their operations [30–32]. Because of their intention to hide among the civilian population, armed forces gain the right to spy on them—as a means of actually protecting real civilians, while preventing terrorists from hiding and planning operations against both civilian and military targets.

If one accepts this line of thought, then potential terrorists [33] deserve the deprivation of their right to privacy (after all they turn against society and pose a serious deadly threat to innocent people); suspects actually benefit from being the objects of surveillance because this way their innocence can be proved and, then, cease to be possible targets of unwanted (and illegal) attacks.

It is quite obvious that according to the above mentioned opinion the consequentialist [34–37] option (depriving the right to privacy so as to prevent a possible —yet really bad or even catastrophic—scenario from happening) would be preferable to the deontological [38–41] approach (doing the right thing, which in our case means that the right of privacy is observed and, even the suspect should only be deprived of it under straightforward circumstances, where there is clear evidence of his/her guilt) because it seems as if the overall benefits [42] (save the lives of innocent people, protect a suspect who proves to be innocent etc.) beat the ethical/legal considerations of whether and under which conditions the right of privacy could be transgressed [43, 44]. In any case, should suspects try to execute planned and co-ordinated attacks against military and political targets, they become fighters and it is, then, legal to treat them as military targets.

From an ethical perspective, one can also comment on these cases from the JWT viewpoint. This tradition has close links with international legislation, especially with reference to the *jus in bello* part, which was fully incorporated into IHL [45]. According to the JWT, for the conduct of war to be just, it must be governed by two principles: those of *discrimination* and *proportionality*. The principle of *discrimination* defines who and what can be justly attacked in a war [46]. It is immoral to deliberately kill non-combatants who are ‘morally and technically innocent, that is, harmless’ [47, 48]; this principle is also included in the IHL where it is clearly stated that non-combatants lives should be protected in the best possible way [49–54]. This principle creates the obligation for military commanders and other people involved in planning a specific military operation to think carefully about the results of the attack; civilians and their property should be protected in the best possible way.

In the case of the war against terrorism, the distinction between combatants and civilians is blurred as a result of the terrorists’ intentional presence in urban areas. If one accepts that surveillance can clear up the grey area between civilians and potential terrorists, then the right of privacy would be rightfully violated. In the cases under consideration, the targets were individuals who were undoubtedly involved in terrorist operations and were high-ranked Al-Qaeda officers. We could, therefore, conclude that they were legitimate targets.

According to the consequentialist argument spying on them made it easier for the principle of discrimination to be observed and civilians were better protected.

Should the right to privacy have prevailed, more people or the wrong people might have been killed; in the end everything was settled down the right way: terrorists were shot dead with the minimal collateral damage. But is this all that matters? Or that should matter? What about the many cases where the right of privacy was suspended but, once the attack was over, no terrorist was killed, just innocent civilians who have been mistaken for suspects or potential threats [55, 56]?

The supplementary principle of *proportionality* is used to resolve issues such as how one should attack and what kind of weapons he might use in order to achieve the military objectives set without causing disproportional collateral damage [57], e.g. civilian casualties or damage to civilian objects. In these non clear-cut cases, where the enemy is surrounded by civilians, and there is a possibility of harming or killing them as a result of the military operation, the principle of double effect is invoked [58]. A high-level approach to this principle is that, if military commanders are planning to perform an act that is likely to harm non-combatants, they can proceed only, if the following four conditions are met [45, 59–61]:

i. The act must be good in itself; it must be a legitimate act of war;
ii. The direct effect must be morally acceptable (the target must be legitimate);
iii. Any negative effect must not be intended;
iv. The intended outcome must be proportional to the foreseen damage. This means that the good effect(s) resulting from the military operation should outweigh any negative consequences of the attack.

Even when the aforementioned conditions are met, still the foreseeable negative consequences must be reduced as far as possible [61].

In both the examples we discussed, there was intelligence about the suspect targets and the attacks were carried out in well chosen locations with the use of weapons that would minimise any side casualties, whilst in both cases the death of high-rank Al Qaeda officers was achieved. The mission was accomplished without casualties for the Americans, and the benefits were important in the case of Yemen (Al-Harithi was killed, the mastermind of many Al Qaeda attacks) and the same holds for Abu Musab al-Zarqawi. Thus, the principle of proportionality was also observed. Surveillance provided crucial information that made the organization and execution of a military mission with confirmed shots and minimal collateral damage possible.

In general, targeted killing attacks aim at weakening the terrorist networks, limit their capacities and prevent their members from planning attacks of a higher magnitude. The tactic of attacking and killing specific targets is superior to other choices for achieving the same results [62] (e.g. the use of troops for engaging into a face to face battle with the terrorists or combining these attacks with air force support) as it does not result in wide aggravating consequences to non-combatants and at the same time protects soldiers from battles that can be avoided. Gathering information on the targets and the area of the operation is extremely important if negative consequences to innocent people and their properties is to be avoided. When information is cross checked from various sources the cost of military operations is reduced not only in financial terms, but in terms of human lives as well [59, 63]. This statement could make a strong utilitarian [64, 65] argument, since the benefits of targeted killing policies extend to non-combatants, as well as military personnel; casualties are reduced in both populations. Beyond that, in the course of war, it is reasonable enough to assume that each opponent organises his/her strategy and decides on the tactical moves in a way that will give him/her the victory having suffered minimal losses [66]; on these terms, the right to life is deemed more important that other rights (to freedom, to privacy etc.) because somebody has to be alive in order to enjoy secondary rights (compared to the fundamental right to life) that the State has to respect, protect and fulfil. If someone is suspect of illegal behaviour and action, then it is reasonable to expect that this person will be deprived of his privacy for the benefit of the innocents he plans to harm. But who will define what kind of actions make somebody a suspect of terrorist activity and when the limit is crossed, so that he/she becomes the object of surveillance [67]? And what might happen to his/her family and friends: they will be deprived of their right to privacy too, as if they are guilty by association? The criteria under which someone poses a terrorist threat are not clear enough [68]. Not only the right to life is deemed important; how life would be like if for the fear of terrorism or unrest other important human rights should be suspended? Definitely it would be a different way of life in many ways compared to what is known to us: a life without freedom, without privacy, without self-determination [69, 70].

According to the second opinion, by analysing the targeted killing attacks, one can argue that fundamental human rights are violated: suspects are killed without been tried and found guilty [71]. Besides that, on the basis of the common article 3 of the Geneva Conventions of 1949, assassinations that are conducted without a prior court decision are illegal. There is a major difference between assassination and targeted killing. When we talk about assassination, we have to distinguish between peacetime assassination and wartime assassination. Peacetime assassination requires all of the following three elements to be present: (1) a murder, (2) of a specific individual, (3) for political purposes. Assassination in wartime has the following characteristics: (1) the specific targeting of a particular individual and (2) the use of treacherous or perfidious means. Other forms of extra-judicial execution, targeted killing, or elimination are not synonymous with assassination. Assassination, whether in peacetime or wartime, constitutes an illegal killing, while targeted killing is the intentional slaying of a specific individual or group of individuals undertaken with explicit governmental approval [72]. Apart from that, a number of *targeted killing* attacks has taken place, where victims were not recognised as terrorists and cases of false intelligence have led to bombing attacks against civilian houses [73]. Beyond the successful operations where it is proved that suspects were indeed involved in terrorist operations, a number of attacks was against civilians and claimed the lives of people whose links to terrorist networks were never proved.

In relation to the Yemen case, it is worth noting that the Amnesty International sent a letter to the president of the United States of America George W. Bush asking him to investigate whether, before the attack, efforts had been made to arrest the terrorist suspects. The letter also asked him to investigate whether the government of Yemen had co-operated in the attack and whether the operation in question was part of a plan for killing suspects rather than attempting to arrest them. The point at which the efforts for arresting suspects of terrorist acts are exhausted is, of course, debatable [74].

Beyond that, the moral dilemmas that are present in all cases of targeted killing are critical. By authorizing the killing of a suspect, it means that his/her right to life is violated. If it is proven afterwards that the suspect was not involved in the acts that he was accused of, then those who carry out the execution have heavy legal and moral responsibilities. On the other hand, if a suspect is free to carry out missions that will lead to mass killings, do those who choose not to confront him/her have moral responsibilities, if nothing else? And how can someone compare the life of a suspect to the life of an innocent civilian? Who is responsible for making such judgments? Is a military commander the suitable judge for these cases? A considerable amount of legal philosophy exists concerning the justification of punishment [75] and who is responsible of deciding on these issues, which can be linked to the targeted killing policies, especially if one considers that the punishment for a suspect terrorist is the death penalty.

These issues are, surely, hard to resolve and anyone can give his own views; however, a minimum level of rights apply, by law, to all suspects and these should be kept. Otherwise, exceptions need to be justified both at the legal level, as well as the moral level, solely on the grounds of the suspects' degree of danger. But could those execptions be justified? It would be difficult for someone to support such a view as fundamental rights—such as been innocent until proven guilty and the right to a fair trial—are been violated. In short, fundamental human rights are been violated just for the fear of future actions of the suspects on the basis of strong evidence, at least as current practice demonstrates [73].

In both our cases, the terrorist suspects were not a direct threat (at the moment of the targeted killing operation) and the 'go ahead' for this kind of operations without a court ruling is a violation of both international humanitarian law and human rights law [76]. On the basis of these ideas the second approach was developed—it maintains that regardless of the benefits targeted killing operations may have, the following consequences should be taken seriously into consideration before initiating such an attack:

- the right of suspects to have a fair trial and the legal principle of been innocent until proven guilty are violated [71]; anyone accused of being a terrorist suspect, even if he/she participated to terrorist operations, should have the right to a fair trial, where the punishment for his/her actions will be decided.
- the ban against arbitrarily taking human lives is violated; the right to life is a fundamental human right and should not be violated in any case, even in situations of national emergency.
- there is a high chance of creating a vicious circle of violence and a chain of terrorist attacks [63].

The focus can then be diverted to the causes that can justify the infringement of rights and the killing of certain individuals [77]. In those cases the same criteria should hold as for pre-emptive/preventive attacks, i.e. one should have strong and well-founded intelligence information and should be absolutely certain that the individual in question is close to launching a terrorist operation against civilians; there should be no other way of preventing the suspect from his/her cause and all other ways of confronting and arresting him/her should have been exhausted and failed [22, 59]. Also, the criteria that make someone a suspect should be clear enough and not just related to his/her ideology, religious beliefs, nationality and other qualities of this type that make whole groups of people altogether suspects.

Concluding this section, in relation to the approaches presented above, I believe that the second is the one that sheds light to the ethics related to the tactic of targeted killing. Neither does it claim that it is the solution to all problems of the war against terrorism, nor does it reject it in all cases—on the contrary, it takes into account all special parameters that characterise every operation, making it different than any previous or next one [78, 79].

In this way, it can not be claimed (or justified) that terrorist suspects should be deprived of their rights all told, become targets and killed in any case, as that would mean that in order to fight terrorism (and manage to keep civilians alive), states and their armed forces should be prepared to accept extended violations of human rights. This might mean that even a policy of "kill now and investigate later" could come in place. In order to avoid extended violations of fundamental human rights, clear rules [80] should apply and a confirmation is necessary before categorizing someone as suspect, spy on him/her, kill him/her on the spot, instead of arresting him/her and providing him/her a with fair trial—a fundamental right of every human being, guilty or innocent [81]. We should not forget that the war against terrorism is, in essence, a battle for protecting the human rights which are violated by the terrorists themselves. Therefore, the states involved in this war, as well as their armed forces, should be the first to show that they respect and protect human rights.

The War Against Terrorism and Privacy: An Impossible Relationship?

A more general question raised in this context is whether the war against terrorism can and needs to be regulated by rules of conduct at all. As frequently stated, since terrorists do not follow the rules, one can fight them without respecting any rule of war either. That stands for the conduct of hostilities, but could one claim it is also true for the protection and respect of human rights [82] of the suspected and prospect terrorists as well? This argument is based on reciprocity: terrorists totally disregard the right to life and other fundamental human rights too (such as the right to freedom, to privacy, to education for the females etc.), so they should be treated accordingly [83]. After all, they started the war, so they are responsible for all the death and destruction that follows [84].

Let's suppose that, in our example, the benefits stemming from surveillance and targeting suspected individuals [85–87] include protecting the innocent civilians and preventing a terrorist attack from happening. Five elements should be taken under consideration:

1. *players*, or decision makers;
2. *strategies* available to each player;
3. *rules* governing players' behavior;
4. *outcomes*, each of which is a result of particular choices made by players at a given point in the game; and
5. *payoffs* accrued by each player as a result of each possible outcome.

We assume that each player (state, terrorists) will pursue the strategies that help him/her to achieve the most profitable outcome in every situation. Each player has a set of preferences for the different possible outcomes and the results of the interaction depend on all the players' decisions. With regards to the topic of spying on suspects for monitoring and collections of information, there are three possible outcomes:

(a) the suspect is innocent and has no connection to terrorism,
(b) the suspect proves to be a terrorist,
(c) the individual under observation has a unclear link to a suspect group and there is no certain answer to whether he could ever be the perpetator of a terrorist act.

On the other hand, the following factors need to be taken into consideration for the agents carrying out the surveillance and planning the targeted killing attack:

- a parameter that would reflect the group in which the suspect is believed to be part of as well as the position of this individual within the group (i.e. a hardliner, a moderate or one of the leaders of the group). This can be represented as: $f(X, Y)$ where X would be representing the group and Y the individual's position. The outcome of this function would be the 'face value' of the information received as an outcome of the surveillance (i.e. $u(I) = f(X, Y)$).
- the ability to use the information gathered by the surveillance, expressed as a function g(I)—it is expected that it should be possible to make use of information with high face value scores (either because the suspect is believed to be part of a 'high-profile' terrorist group or because he/she is believed to be high in the hierarchy of the group)—this would mean that $\frac{\partial f(X,Y)}{\partial X} > 0$ and $\frac{\partial f(X,Y)}{\partial Y} > 0$.
- the ability to act with (or without) the information received from spying on the suspect and, thus, disrespecting his right to privacy (while he/she has not proved to be guilty). This can be expressed as a function $a_{\text{with}}(I) = (1 + \delta) * a_{\text{without}}(I)$ where δ is an adjustment factor to cater for the loss (or presence of the information) and $a_{\text{without}}(I)$, $a_{\text{with}}(I)$ represent the cost of acting without (respectively: with) the information to be acquired by spying on the suspect.
- the positive (or negative) results of using the information; of course, in this case we need to take into account the credibility of the information received. In this case, the results from the use of information are calculated as: $r_{\text{use}}(I) = \kappa * r(I) * P(I \mid X, Y)$ where κ is an adjustment factor for the importance of the information

(positive if advancements are made, negative if it might lead into more difficult situations), r(I) is a function describing the expected results from the use of the information and P(I | X, Y) is the conditional probability that the information is true, given the individual who is the source and the group that he is believed to belong to.

- the ability to get the same information in a different way without limiting the suspect's rights. This would be represented by a function a(I) = c(I) * c(I, Z) where c(I) is the current cost of getting the information and c(I, Z) is a function representing the magnitude of difference if information I is achieved in way Z.
- the cost of getting the information requested in a way that observes international legislation. This cost will be calculated on the basis of the sub-costs for delaying mining the information because of following international legislation procedures, as well as a sub-cost that would include the reaction to the public for this effort. We can write this as $c_{IL}(I) = c(I) * c(d(I), r(I))$ where c(I) is the current cost of getting the information and c is a function that would take into account whether this will slow down the process (and, thus, create security problems) as well as the reaction of the public to such changes.
- the cost of getting the information using methods that are in line with human rights legislation. This is represented in the form of $c_{HR}(I) = \lambda * c(I)$, where λ is an adjustment factor to count for the change in the surveillance process.
- the cost of the image of the state in the international scene—this is represented as c_{IS} and is equal to $c_{IS}(I) = (1 + \mu) * c(I)$ where c(I) represents the current cost of getting the information and μ is an adjustment factor that shows the change (for better or worse) to the image of the state in the international scene after the information was acquired and acted upon. We expect this function to be monotonic with regards to the information variable, so $\frac{\partial c_{IS}(I)}{\partial(I)} > 0$.
- the short/long-term results of the current mining methods expressed as a cost. This can be modelled on the basis of the magnitude that current methods will have on acquiring information in the short and long-term, so we need to scale the current information cost function to match the predictions. This is modelled as $c_{SLT} = (1 + \theta) * c(I)$ where θ models the adjustment factor for any changes in the cost for information mining.

In order to decide whether surveillance gains should exceed costs (because of the violation of human rights) we have to calculate the result provided by the utility function for the expected information added to the function giving the results from the use of information should exceed all other costs added up together. In other words, the following inequality will need to hold:

$$u(I) + \mathrm{r}_{\mathrm{use}}(\mathrm{I}) > \mathrm{g}(\mathrm{I}) + \mathrm{a}_{\mathrm{with}}(\mathrm{I}) + \mathrm{a}(\mathrm{I}) + \mathrm{c}_{\mathrm{IL}}(\mathrm{I}) + \mathrm{c}_{\mathrm{HR}}(\mathrm{I}) + \mathrm{c}_{\mathrm{IS}}(\mathrm{I}) + \mathrm{c}_{\mathrm{SLT}} \tag{1}$$

If the suspect proves to be innocent (outcome a), then the state/military has the benefit of using the information gained in order to actually protect him (he/she will not be assaulted by targeted killing). Of course, the means of gathering the data contradicts human rights conventions [88, 89], so there are negative results in relation to the state's image to international community; also, there can both short and long-term results because of the illegal mining methods that could upset both terrorists (vicious cycle of violence) but also informed citizens who oppose to governmental illegal practices no matter the cause.

If the suspect proves to be guilty (outcome b), then the state/military has the benefit of using the information gained in order to actually protect the society (he/she will be eliminated by targeted killing) with minimum collateral damage [90]. Of course, the means of gathering the data still contradicts human rights conventions, so there are negative results in relation to the state's image to international community; also, there can both short and long-term results because of the illegal mining methods that could further upset both terrorists (vicious cycle of violence) but also informed citizens who oppose to governmental illegal practices no matter the outcome of the investigation.

If the individual under observation is vaguely linked to a suspect group (outcome c), then further questions arise as to how he/she should be treated. The issue of Intel analysis and surveillance of specific individuals, deemed (or marked out due to ideological or religious beliefs) as suspects of potentially planning or executing a dangerous act against the common good, will definitely come into the foreground [91, 92]. For many people, under the terrorism threat, any further infringement of a suspect's right to privacy could be justified; intruding into his/her life is considered/regarded as the morally right thing to do and a governmental duty of the utmost importance in order to prevent bad things from happening and, thus, to protect civilians from an unjust terrorist attack. But could (or should) that practice ever be justified? That, practically, could mean that one is allowed to spy on preventive grounds and even attack suspected terrorists directly or use prohibited methods [93] to keep the pressure on anyone, in order to make the terrorists come out.

This notion is problematic, as it makes the state's obligation to protect human rights (among them the right to privacy) and observe the relevant national and international laws and regulations optional, depending on the enemy one faces [94, 95]; but when a state ratifies a treaty, whether it concerns the rights of civilians or the prohibition of certain weapons, it has to adhere to it. If the enemy has not ratified the same treaties or shows no respect for the Law and probably chooses to use prohibited methods of war, this is not an excuse for the armed forces to act in the same way. States have to adhere to both national regulations and the treaties they ratified; this is not only an obligation of the United States, but rather an obligation of the international community [96]. This is the reason why the armed forces have to do everything in their power to defeat the enemy using tactics that are in accordance with IHL.

As far as the right of privacy is concerned, when the armed forces can spy on any civilian they believe has a link to terrorism (because he/she may visit certain sites on internet, has ideological or religious believes that could turn him/her into a terrorist etc.) based on a preventive (a terrorist attack will be stopped before being executed) or a vague utilitarian (many people will be saved) argumentation, then other fundamental rights are effected as well. Questions concerning civilians' freedom of action, communication, expression arise; questions concerning the equality of people living in different countries or adhering to different religious traditions may arise too: for example are human rights of an Iraqi citizen protected and respected the same way that human rights of an American citizen are preserved? Is there any excuse for violating the right to privacy of whole populations who live in a specific area or use dubious phraseology in their online communications? One must draw a limit, well and clearly defined, which, when crossed, would provide the reason behind civilians' and suspects' surveillance. Otherwise, everyone could become a suspect for the pettiest reasons and be deprived of his established—by law—rights.

Another important issue that is raised as a result of the targeted killing operations is the role the American Government and US military forces have to play in the war against terrorism [97], which is a war against a 'stateless terrorist enemy'. The surveillance of suspects and the attacks against individuals is purely of pre-emptive, even preventive nature, as the primary target of these operations is the death of high-rank terrorists in order to prevent, avoid or—at least—reduce the numbers of future terrorist attacks [98]. Furthermore, surveillance and attacks of this kind primarily aim not only at eliminating terrorists but also target those who are offering them operational, military or any other kind of support [99, 100].

More specifically, the United States have relied on covert action in their campaign against terrorism and it appears that surveillance and targeted killing attacks run in co-operation with other governments, in full secrecy if needed [101]. Both the tactic of spying on civilians using various means and methodologies and the related targeted killing attacks may be carried out with or without the consent of those governments whose authorities are involved. This kind of reasoning can become very dangerous as the US Government and military forces will be able to act as an 'international police force', responding to any new security challenges unilatery and even pre-emptively [102] (The consent of the state concerned can be linked to discussions on the evolving notion of state sovereignty [103]). That scenario gives them full power to interfere and intervene wherever they suspect there is on-going terrorist activity violating established human rights even before entering the real battlefield. A number of related interesting issues also arise from the Yemen incident; one of these, which due to limited space we can not fully analyze here, is that the US military forces took military action outside their borders.

If one looks at this issue from the viewpoint of international legislation, he will see that surveillance and the conduct of military operations (for any reason) on other nation's territory are problematic, especially if they do not have the consent of the home government. The United Nations Charter unambiguously states the cases where an operation is permitted: an approval is needed either from the Security Council or a peripheral organisation, or the military action should be justified on the basis of exercising the legal right of states to self-defence, either individually or collectively [104]. It becomes even more complex and hard to justify it legally when someone realises that the operation on foreign territory happens on the suspicion of terrorist act. In this case any principle of sovereignty is overlooked [105–107]. Since President Donald Trump gave the Central Intelligence Agency

(CIA) new authority to conduct drone attacks against suspected militants (a few months ago) [108], the discussion of this issue has become of extreme importance.

The justification of the above mentioned line of reasoning is problematic on moral grounds as well. Even if arguments about ending the war sooner with less side casualties are used, they remain in essence unconvincing. The effort to justify extensive wide-ranging domestic and international political spying on people that in certain cases could end up in targeted killing operations on the basis that they are been carried out for securing world peace that is in danger from terrorists is not fully acceptable, as fundamental rights—also part of international legislation—are been violated [109].

With respect to the benefits arising from government mass surveillance and the following military actions, one should scrutinize them carefully, as the results of the American policy will be visible in the years to come. The US have turned out to be the most powerful country in the New World Order and their contribution so far may indeed be considered as vital, although one has to keep in mind that the great power they hold should not suppress other states' rights or violate human rights; the power to use military force against any possible suspected enemy could provoke counteractions and become a major source of international disputes leading to infringement of human rights and breaches of peace.

Thus, it is clear that the way US military and its allies fight this war—as well as the way it treats suspects, terrorists and their organizations—is crucial as it will form a predecessor for future operations. It is very important that the latest technology (especially weapons with a high degree of accuracy) is employed in order to minimize civilian casualties instead of applying extensive mass electronic surveillance against practically everybody, monitoring and gathering of data, targeting certain "trigger" words or phrases and producing suspects on the grounds of visiting certain types of web sites, or communicating via email or online chat with suspicious individuals or groups. Definitely, there are benefits when surveillance is used wisely, however, in the case of monitoring and spying on suspects who are then targeted, one should be very careful; in the cases under discussion, the attack was indeed launched against terrorists. In the past US forces spied on and then attacked groups of people whom they mistakenly took for terrorists: as an example one could refer to the accidental bombing of an Afghan wedding party which was considered to be a Taliban gathering [73]. Therefore, the operations for locating and targeting suspects should be carried out in such a way that will ensure the people targeted are the right ones and even so, eliminate the possibility of other people being in the vicinity and getting hurt.

Besides the practical aspect of this matter, it follows that a major theoretical issue that needs to be resolved is whether ethical principles, especially those governing the protection of human rights and the conduct of war, are applicable in the 'war' against terrorism. Many people would argue that, the principles of discrimination and proportionality can not be applied because terrorists get intentionally mixed up with civilians [110]. This tactic results in US forces efforts to avoid civilian casualties being futile, while at the same time methods such as mass monitoring of people's behavior, activities, or other information for the purpose of identifying potential terrorists

[111, 112] and protecting the innocent, seems to produce negative effects and a cycle of violence associated with acts of retribution and revenge, that, in the end, only make the war last longer, leading to more casualties [73].

Concluding Remarks

It is true that the war against terrorism is a new type of war: it takes place in towns, villages and urban areas; there is no 'visible' enemy to fight; terrorist tactics are ever-evolving; they intentionally blend in with civilians and use non-combatants either to form a 'human shield' around them or to harm them as means of making their voice heard and their demands met. All these make states and militaries involved in the fight want to bring the war to an end as quickly as possible [73], but should everything towards this goal be excused or permitted [113]?

Definitely not, every new challenge should not become a reason for abandoning or modifying fundamental moral values in order to be in line with any temporal interests [114]. This approach might prove a dangerous slippery slope undermining the effort of our society to honor individuals rights and to discriminate between innocent/guilty, non-combatants/combatants. The distinction between right/wrong, moral/immoral, legal/illegal is rather important; laws and regulations that protect human rights during peace and war should be observed because otherwise, if the enemy defines our actions to repel him, there would be no stable moral or legal system to refer to. No matter the consequences, even when in extremely critical situations, one should hold true to his moral and legal obligation to do what is right, thus to protect the established rights of the individuals both during peace and in wartime [115, 116].

In the case of targeted killing [117], even before the formulation of IHL regarding the conduct of war, there were moral principles governing what was acceptable and what was not: Plato, Aristotle and Cicero grant special status to non-combatants. Under which specific circumstances can a suspected individual be deemed as a combatant, thus loose his rights and become a target of surveillance—and maybe of attack as well [118]? And can someone become the object of government monitoring during peace and without having executed any illegal action [119, 120]? If an affirmative answer is given to this question, it could then be deemed as acceptable and permissible to disrespect suspected civilians' right to privacy and limit their freedom on pre-emptive, even preventive, grounds and with an obscure intention of protecting the real innocents. Also, when potential enemies are identified among different populations and nationalities (in the war against terrorism example, they could be a part of the US population, the Iraqi people etc.): how one should deal with them (i.e. do they have the same rights despite their different nationality) [121]? How many should one expect to save in order to justify sacrificing another person's rights? And are there really clear established criteria so as to detect a potential terrorist? Where should someone draw the line between the status of an innocent person and a suspect (who might not be innocent beyond reasonable doubt, but is not obviously guilty too)? And who is to decide on how to treat the suspect? Who is the enemy and when does the war really start?

The just conduct of war as analyzed in depth by St. Augustine [122] and Hugo Grotius [123] may not be enough for the contemporary military operations. The evolution of just war principles though morally ground IHL [124, 125]. Therefore, the development of the existing legislation was carried out on moral grounds and principles concerning war. And the evolution continues so as to cover new forms of military operations and new types of enemies [126]. Abolishing the rules and the obligations stemming from the treaties and the conventions a state has ratified is like giving away its moral values.

So, on the contrary, one, instead of lowering his moral standards, should seek alternative ways of achieving his goals without having to follow an immoral path. As a result, any extension of the capabilities of military technology and, consequently, of military capacity should not make a state give away its moral values. Instead, the most current and accurate weapons should be used in order to fulfil the moral and legal obligations, as these are derived from IHL and human rights conventions.

References

1. Dinstein Y (2001) War, aggression, and self-defense. Cambridge University Press, Cambridge
2. Franck T (2004) Recourse to force: state action against threats and armed attacks. Cambridge University Press, Cambridge
3. Kegley CW (2003) The New global terrorism: characteristics, causes, controls. Prentice Hall, New Jersey
4. Cryer R (2002) The fine art of friendship: Jus in Bello In Afghanistan. J Conflict Secur Law 7(1):37–83
5. Sassòli M (2006) Terrorism and war. J Int Crim Just 6:959–981
6. Meggle G (2004) Terror and counter-terror: initial ethical reflections. In: Textor M, Kemmerling A, Meggle G (eds) Ethics of terrorism and counter-terrorism. De Gruyter, Berlin, pp 161–176
7. Thompson J (2004) Terrorism, morality and right authority. In: Textor M, Kemmerling A, Meggle G (eds) Ethics of terrorism and counter-terrorism. De Gruyter, Berlin, pp 151–160
8. Shue H (2008) Do we need a 'morality of war'? In: Rodin D, Shue H (eds) Just and unjust warriors: the moral and legal status of soldiers. Oxford University Press, Oxford, pp 87–111
9. Rodin D (2011) Morality and law in war. In: Strachan H, Scheipers S (eds) The changing character of war. Oxford University Press, Oxford, pp 446–463
10. Blum G, Heymann P (2010) Law and policy of targeted killing. Harv Natl Secur J 1:145–170
11. Schmitt M (2011) Drone attacks under the Jus ad Bellum And Jus in Bello: clearing the 'Fog of Law'. Yearb Int Humanit Law 13:311–326
12. Lyon D (2002) Surveillance as social sorting: privacy, risk and automated discrimination. Routledge, Abingdon/New York
13. Ball K, Haggerty K, Lyon D (2012) Handbook of surveillance studies. Routledge, Abingdon/New York

14. Leone RC, Anrig GJ (2003) The war on our freedoms: civil liberties in an age of terrorism. The Century Foundation, New York
15. Torpey J (2007) Through thick and thin: surveillance after 9/11. Contemp Sociol 36(2): 116–119
16. Monahan T (2010) Surveillance in the time of insecurity. Rutgers University Press, New Brunswick
17. Taurek JM (1977) Should the numbers count? Philos Public Aff 6(4):293–316
18. Patterson E, Casale T (2005) Targeting terror: the ethical and practical implications of targeted killing. Int J Intell Counter Intell 18(4):638–652
19. Ben-Naftali O, Michaeli KR (2003) Justice-ability: a critique of the alleged non-justifiability of Israel's policy of targeted killings. J Int Crim Just 1(2):368–405
20. Fitzpatrick J (2003) Speaking law to power: the war against terrorism and human rights. Eur J Int Law 2:241–264
21. Teplitz RF (1995) Taking assassination attempts seriously: did the United States violate international law in forcefully responding to the Iraqi plot to kill George Bush? Cornell Int Law J 28(2):569, 610–613
22. Kretzmer D (2005) Targeted killing of suspected terrorists: extra-judicial executions or legitimate means of defence? Eur J Int Law 16(2):171–212
23. Turns D (2017) The United Kingdom, unmanned aerial vehicles, and targeted killings. ASIL 21(3). https://www.asil.org/insights/volume/21/issue/3/united-kingdom-unmanned-aerial-vehicles-and-targeted-killings. Accessed 15 June 2017
24. Gray C (2016) Targeted killing outside armed conflict: a new departure for the UK? J Use Force Int Law 3(2):198–204
25. Colonomos A (2017) The global targeted killings bandwagon: who's next after France? The conversation (February 8, 2017). http://theconversation.com/the-global-targeted-killings-bandwagon-whos-next-after-france-71840 Accessed 15 June 2017
26. Werner W (2015) Targeted and non-targeted killing. De Ethica J Philos Theol Appl Ethics 2 (1):49–60
27. Addicott J (2002) The Yemen attack: illegal assassination or lawful killing? http://jurist.law.pitt.edu/forum/forumnew68.php. Accessed 25 June 2017
28. Lekea JK (2003) 'Missile Strike Carried Out With Yemeni Cooperation'—the war against terrorism: a different kind of war?. J Mil Ethics 2(3):230–239
29. Pincus W (2003) Missile strike carried out with Yemeni cooperation: official says operation authorized under bush finding. J Mil Ethics 2(3):227–229
30. Convention (II) for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea, Article 13
31. Geneva Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, Article 13
32. Convention (III) relative to the Treatment of Prisoners of War of August, 12 1949, Article 4
33. David S (2003) If not combatants, certainly not civilians. Ethics Int Aff 17(1):138–140
34. Buchanan A (2010) Human rights, legitimacy, and the use of force. Oxford University Press, Oxford, pp 281–282
35. Cushman T (2005) The human rights case for the war in Iraq: a consequentialist view. Cambridge University Press, Cambridge, pp 78–107
36. Fabre C (2009) Guns, food, and liability to attack in war. Ethics 120(1):36–63
37. McMahan J (2008) The morality of war and the law of war. In: Rodin D, Shue H (eds) Just and unjust warriors: the moral and legal status of soldiers. Oxford University Press, Oxford, pp 19–43
38. Dougherty T (2013) Rational numbers: a non-consequentialist explanation of why you should save the many and not the few. Philos Q 63:413–427
39. Halstead J (2016) The numbers always count. Ethics 126:789–802
40. Moore M (2008) Patrolling the borders of consequentialist justifications: the scope of agent-relative obligations. Law Philos 27(1):35–96

41. Moore MS (2012) Targeted killings and the morality of hard choices. In: Finkelstein C, Ohlin JD, Altman A (eds) Targeted killings: law and morality in an asymmetrical world. Oxford University Press Oxford, pp 434–466
42. Manning P (2008) A view of surveillance. In: Leman-Langlois S (ed) Technocrime: technology, crime, and social control. Willan Publishing, Devon, UK, pp 209–242
43. Treverton GF (2003) Terrorism, intelligence and law enforcement: learning the right lessons. Intell Natl Secur 18(4):121
44. Ball K, Webster F (eds) (2003) Intensification of surveillance. Crime, terrorism and warfare in the information age. Pluto Press, London
45. Johnson JT (1999) Morality and contemporary warfare. Yale University Press, New Haven, CN and London, pp 8–40
46. Nicholas F (1996) Who, what, when and how to attack. http://atlas.usafa.af.mil/jscope/JSCOPE96/fotion96.html. Accessed 25 Sept 2016
47. Mapel DR (1996) Realism and the ethics of war and peace. In: Nardin T (ed) The ethics of war and peace. Secular and religious perspectives 67. Princeton University Press, Princeton, NJ
48. Holmes RL (1989) On war and morality. Princeton University Press, Princeton, NJ, p 104
49. Convention (IV) relative to the Protection of Civilian Persons in Time of War, Articles 4, 27 3/34
50. Protocol Additional to the Geneva Conventions I, Article 52
51. Common Article 3 of the 1949 Geneva Conventions. Protocol Additional to the Geneva Conventions I, Article 48
52. Geneva Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, Articles 19, 21, 24, 25
53. Convention (II) for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea, Articles 22–35
54. Protocol Additional to the Geneva Conventions I, Articles 12 and 23
55. O'Connell ME (2012) Unlawful killing with combat drones: a case study of Pakistan, 2004–2009. In: Bronitt S, Gani M, Hufnagel S (eds) Shooting to kill, the law governing lethal force in context. Hart Publishing Ltd., Oxford, OX, pp 263–291
56. Farhat T (2010) The year of the drone misinformation. Small Wars Insurg 21(3):529–535
57. Schmitt M (1992) State-sponsored assassination in international and domestic law. Yale J Int Law 17:609–685
58. Cavanaugh TA (2006) Double-effect reasoning: doing good and avoiding evil. Clarendon Press, Oxford
59. Gross E (2002) Self-defense against terrorism-what does it mean? The Israeli perspective. J Mil Ethics 1(2):105
60. Regan RJ (1999) Just war, principles and cases. The Catholic University of America Press, Washington D.C., pp 95–96
61. Walzer M (1984) Just and unjust wars, a moral argument with historical illustration. Basic Books, Harmondsworth, Middlesex p 153
62. Clough B (2002) Unmanned aerial vehicles: autonomous control challenges, a researcher's perspective. In: Murphy R, Pardalos PM (eds) Cooperative control and optimization. Kluwer, Dordrecht, pp 35–52
63. Statman D (2005) Targeted killing. In: Shanahan T (ed) Philosophy 9/11. Thinking about the war on terrorism 183-202. Open Court Publishing Company, Chicago
64. Shaw WH (2016) Utilitarianism and the ethics of war. Routledge, New York
65. Shaw WH (2014) Utilitarianism and the ethics of war. In: Eggleston B, Miller DE (eds) The Cambridge companion to utilitarianism. Cambridge University Press, Cambridge, pp 303–324
66. Cohen A, Shany Y (2007) A development of modest proportions. The application of the principle of proportionality in the targeted killings case. J Int Crim Just 5:310–321

67. Fenwick H (2010) Recalibrating ECHR rights, and the role of the human rights act post 9/11: reasserting international human rights norms in the war on terror? In: Letsas G, O'Cinneide C (eds) Current legal problems. Oxford University Press, Oxford, pp 153–243
68. Cragin K, Sara AD (2004) The dynamic terrorist threat. RAND, Santa Monica, CA. https://www.rand.org/content/dam/rand/pubs/monograph_reports/2005/MR1782.pdf. Accessed 15 June 2017
69. Harcourt B (2007) Against prediction. Profiling, policing, and punishing in an actuarial age. University of Chicago Press, Chicago
70. Leenes R, Koops EJ, De Hert P (eds) (2008) Constitutional rights and new technologies. A comparative study. T.M.C. Asser Press, The Hague (Information Technology & Law Series, vol 15)
71. von Schorlemer S (2003) Human rights: substantive and institutional implications of the war against terrorism. Eur J Int Law 2:265–282
72. Harder TJ (2002) Time to repeal the assassination ban of executive order 12,333: a small step in clarifying current law. Mil Law Rev 172:1–39
73. Pfaff T (2003) Non-combatant immunity and the war on terrorism. http://www.usafa.af.mil/jscope/JSCOPE03/Pfaff03.html. Accessed 15 June 2017
74. There have been efforts, of course, to give strict instructions about when the use of lethal force is allowed. In the case of McCann and Others v. the United Kingdom (European Court of Human Rights) the judgment gives very specific criteria for the use of lethal force. For more information, see Cohen, Amichai and Yuval Shany, *supra*. Also, see Sperotto, Federico, 2007. Violations of Human Rights during Military Operations in Chechnya. Working paper no. 41. http://www.du.edu/gsis/hrhw/working/2007/41-sperotto-2007.pdf. Accessed 15 June 2017
75. Wood D (2002) Retribution, crime reduction and the justification of punishment. Oxford J Legal Stud 22(2):301–321
76. Amnesty International (2002) Yemen/USA: government must not sanction extra-judicial executions. http://web.amnesty.org/library/index/EN-GAMR511682002. Accessed 15 June 2017
77. Welsh BC, Farrington DP (2009) Making public places safer surveillance and crime prevention. Oxford University Press, Oxford
78. Alston P (2011) The CIA and targeted killings beyond borders. Harv Natl Secur J 2(2):283–446
79. Barnidge RP Jr (2012) A qualified defense of American drone attacks in northwest Pakistan under international humanitarian law. Boston Univ Int Law J 30(2):410–447
80. Davis LE, McNemey M, Greenberg MD (2016) Clarifying the rules for targeted killing. An analytical framework for policies involving long-range armed drones. RAND Corporation, Santa Monica, CA
81. Downes C (2004) 'Targeted killings' in an age of terror: the legality of the Yemen strike. Journal of Conflict and Security Law 9(2):277–294
82. Schmitt MN (2012) Unmanned combat aircraft systems and international humanitarian law: simplifying the often benighted debate. Boston Univ Int Law J 30(2):595–619
83. Klabbers I (2003) Rebel with a cause? Terrorists and humanitarian law. Eur J Int Law 2:311
84. Roberts A (2002) Counter-terrorism, armed force and the laws of war. Survival 44(1):10
85. Margalit A (2012) Did LOAC take the lead? Reassessing Israel's targeted killing of Salah Shehadeh and the subsequent calls for criminal accountability. J Conflict Secur Law 17 (1):147–173
86. Jahagirdar OM (2008) Targeted killing, not assassination: the legal case for the United States to kill terrorist leaders. J Islamic Law Cult 10(2):234–251
87. Murphy R, Radsan AJ (2009) Due process and targeted killing of terrorists. Cardozo Law Rev 31(2):405–450
88. Parent WA (1983) Privacy, morality and the law. Philos Public Aff 12(4):269–288
89. Allen AL (2008) The virtuous spy: privacy as an ethical limit. The Monist 91(1):3–22
90. McNeal GS (2014) Targeted killing and accountability. Georgetown Law J 102:681–794

91. Tavani HT, Moor JH (2001) Privacy protection, control of information, and privacy-enhancing technologies. Comput Soc 31(1):6–11
92. Reiman J (2004) Driving to the panopticon: a philosophical exploration of the risks to privacy posted by the information technology of the future. In: Rössler B (ed) Privacies: philosophical evaluations. Stanford University Press, Stanford, pp 194–214
93. Protocol Additional to the Geneva Conventions I, Article 51
94. Dutta N (2004) The face of the other. Terror and the return of binaris. Interventions 6 (3):431–450
95. Cole D (2005) Enemy aliens: double standards and constitutional freedoms in the war on terrorism. The New Press, New York
96. Condorelli L, de Chazournes LB (1984) Quelques remarques a propos de l'obligation des Etats de 'respecter et faire respecter' le droit international humanitaire en toutes circonstances. In: Swinarski C (ed) Etudes et essais sur le droit international humanitaire et sur les principes de la Croix-Rouge en l'honneur de Jean Pictet. Springer, New York, pp 17–35
97. Tinetti J (2004) Lawful targeted killing or assassination? A roadmap for operators in the global war on terror. Naval War College, Newport, RI
98. Calhoun L (2003) The strange case of summary execution by predator drone. Peace Rev: A J Soc Just 15(2):209–214
99. Blank LR (2012) Targeted strikes: the consequences of blurring the armed conflict and self-defense justifications. William Mitchell Law Rev 38(5):1655–1700
100. Guiora AN (2009) Not "by all means necessary": a comparative framework for post-9/11 approaches to counterterrorism. Case West Reserve J Int Law 42(1&2):273–287
101. Murphy SD (2003) International law, the United States, and the non-military "War" on terrorism. Eur J Int Law 2:363
102. Benvenisti E (2004) The US and the use of force: double-edged hegemony and the management of global emergencies. Eur J Int Law 15(4):677–700
103. Lake DA (2003) The new sovereignty in international relations. Int Stud Rev 5(3):303–323
104. UN Charter, Chapter VII, Articles 39, 42, 51–54
105. Lee S (2006) International Governance and the Fight against Terrorism. Ethics and International Affairs 20(2):241–246
106. Cohen JL (2004) Whose sovereignty? Empire versus international law. Ethics Int Aff 18 (3):1–24
107. Bartelson J (2006) The concept of sovereignty revisited. Eur J Int Law 17(2):463–474
108. The Independent, 14 March 2017. http://www.independent.co.uk/news/world/americas/donald-trump-cia-power-drone-strikes-military-a7628561.html. Accessed 15 June 2017
109. Lucas GR Jr (2003) The role of 'international community' in just war tradition—confronting the challenges of humanitarian intervention and preemptive war. J Mil Ethics 2(2):122–144
110. Zupan DS (2006) Just war theory, law enforcement and terrorism: a reflective equilibrium. http://www.usafa.af.mil/jscope/JSCOPE03/Zupan03.html. Accessed 15 June 2017
111. Gross E (2002) The influence of terrorist attacks on human rights in the United States: the aftermath of September 11, 2001. North Carolina J Int Law Commer Regul 28(1):1–102
112. Copeland RA (2004) War on terrorism or war on constitutional rights? Blurring the lines of intelligence gathering in post-September 11 America. Texas Tech Law Rev 35:1–31
113. Banks WC (ed) (2011) New battlefields, old laws: critical debates on asymmetric warfare. Columbia University Press, New York
114. Galloway HH (2002) Don't forget what we're fighting for: will the fourth amendment be a casualty of the war on terror? Wash Lee Law Rev 59(3):921–974
115. Barela SJ (2014) International law, new diplomacy and counterterrorism: an interdisciplinary study of legitimacy. Routledge, Abingdon
116. Blank LR, Noone GP (2013) International law and armed conflict: fundamental principles and contemporary challenges in the law of war. Wolters Kluwer Law & Business, New York
117. Bergen PL, Rothenberg D (eds) (2015) Drone wars: transforming conflict, law, and policy. Cambridge University Press, New York

118. Merola LM (2012) Evaluating the legal challenges and effects of counterterrorism policy. In: Lum C, Kennedy LW (eds) Evidence-based counterterrorism policy. Springer, New York, pp 281–300
119. Hicks N, McClintock M (2004) Defending security: the right to defend rights in an age of terrorism. Human Rights First, New York
120. Cole D, Dempsey JX, Goldberg C (2002) Terrorism and the constitution: sacrificing civil liberties in the name of national security. New Press, New York
121. Sekhon V (2003) The civil rights of "others": antiterrorism, the patriot act, and Arab and South Asian American rights in post-9/11 American Society. Texas J Civil Libert Civil Rights 8(1):117
122. Ramsey P (1992) The just war according to St. Augustine. In: Elshtain JB (ED) Just war theory. New York University Press, New York, pp 8–22
123. Christopher P (1999) The ethics of war and peace: an introduction to legal and moral issues. Pearson, Upper Saddle River, New Jersey, pp 81–103
124. O'Brien, WV (1981) The conduct of just and limited war. Praeger, New York, pp 37–70
125. Westhusing T (2003) Taking terrorism and ROE seriously. J Mil Ethics 2(3):3
126. Finnane MJC, Donkin S (2013) Fighting terror with law? Some other genealogies of pre-emption. Int J Crime Just Soc Democr 2(1): 3–17

Part III
Cyber Surveillance

Chapter 15
Data Hiding in the Wild: Where Computational Intelligence Meets Digital Forensics

Victor Pomponiu, Davide Cavagnino and Marco Botta

Abstract In the context of an increasing dependence on multimedia contents, data hiding techniques, such as watermarking and steganography, are becoming more and more important. Due to the complementary nature of their general requirements, i.e., imperceptibility, robustness, security and capacity, many data hiding schemes endeavour to find the optimal performances by applying various approaches inspired from nature. In this paper, we provide a review and analysis of the main computational intelligence approaches, including Artificial Neural Networks (ANNs) and Fuzzy Sets (FSs), which are employed in information hiding. Furthermore, with the aid of the recent state of the art, we discuss the main challenges to be addressed and future directions of research.

Keywords Computational intelligence · Artificial neural networks
Fuzzy sets · Information hiding · Security · Surveillance

Introduction

Due to the widespread distribution of digital media and development of sophisticated multimedia manipulation tools, data hiding, which aims to protect the intellectual property of media contents, has become an important research area, both in industry and academic institutions [1, 2].

Depending on the characteristics of the embedded information, data hiding systems can be split into two broad techniques: *steganography* and *watermarking*. The former technique establishes a covert communication channel between two

V. Pomponiu (✉)
Agency for Science, Technology and Research, 1 Fusionopolis Way,
487372 Singapore, Singapore
e-mail: victor.pomponiu@ieee.org

D. Cavagnino · M. Botta
Computer Science Department, University of Torino, C. Svizzera 185,
CAP 10149 Turin, Italy

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_15

parties and is characterized by concealing the existence of the secret message hidden within an innocuous content. Instead, watermarking consists of embedding secret information in a digital content, e.g., audio signal, image, and video, with security purposes. Unlike to steganography, the existence of the watermark is not secret and it is related to the content in which it is inserted [3].

The secret information can be embedded in *spatial domain*, *transform domain*, or *compressed domain*. In case of digital images, a well-known embedding strategy resides in hiding the secret data directly into distributed pixels. The Least Significant Bit (LSB), the Correlation Based, the CDMA Spread Spectrum and the Patchwork are representative approaches of this domain [4]. However, performing the insertion in spatial domain raises several major issues, e.g., vulnerability against content manipulations.

On the other hand, in the transform domain the content is first changed into transform coefficients, and then the resulted coefficients are modified in order to embed the secret data. Some of the most common schemes in this domain are:

- those related to frequency domain, such as the Discrete Fourier Transform (DFT) [5], the Discrete Cosine Transform (DCT) [6], the Discrete Wavelet Transform (DWT) [7], the Karhunen-Loève Transform (KLT) [8] and the Slant Transform [9];
- those operating in fractal domain [10]
- techniques based on Singular Value Decomposition (SVD) [11, 12].

Steganography and watermarking have many characteristics in common. Nevertheless, they differ in several ways such as scope, requirements and extraction/detection of the secret data (Table 15.1). Unlikely from steganography, since the presence of the watermark is known, the watermarking has two crucial requirements: robustness against content manipulation and security attacks. Instead, the essential requisites for steganography are perceptual invisibility and statistic undetectability.

Table 15.1 Steganography and watermarking: scope, requirements and type of detection/extraction

Technique	Scope	Requirements	Detection/Extraction
Steganography	Covert communication	Perceptual invisibility Statistical invisibility Large data payload	Non-blind[a] Semi-blind[b] Blind[c]
Watermarking	Copyright protection Content authentication Transaction tracking Copy control	Imperceptibility Robustness Security	

[a] *Non-blind* by using at least the original content

[b] *Semi-blind* by using information related to the original content

[c] *Blind* without resorting the original content

The main difficulty in creating reliable data hiding techniques is caused by the complementary nature of their requirements which impose several tradeoffs. For instance, embedding large amount of data degrades the visual quality of the media content, while increasing the imperceptibility of the secret data lowers its robustness against attacks. Therefore, in the last years studies have been conducted to devise optimal detectors to improve the detection of the secret data in case of attacks [13–18]. Furthermore, to increase the imperceptibility of the hidden data, various perceptual models that exploit the characteristics of the human visual and auditory system have been proposed [19].

Computational Intelligence (CI) is a well-defined branch of research which exploits different levels of imprecision to devise computational paradigms that leads to acceptable solutions to hard problems, i.e., NP-complete problems, for which an exact solution cannot be derived in a reasonable time [20, 21]. Recently, computational intelligence techniques have received considerable attention, and many approaches have been proposed to improve the performances of the data hiding schemes. It can be clearly observed that the number of papers using computational intelligence to devise data hiding schemes has increased, imposing an urgent analysis of the current situation.

In this paper, we summarize and organize research results in the field which combines computational intelligence and multimedia security in a novel way that integrates and adds understanding in order to provide the reader with a clear overview. Consequently, we classify the existing literature developing a perspective on the area, besides evaluating the new trend directions.

To provide a better understanding of how the CI technologies are employed in data hiding, we organize the rest of the chapter in five sections: section "Background'' presents the most important properties of the main soft computing technologies applied in data hiding such as Artificial Neural Networks (ANNs) and Fuzzy Sets (FSs) and further defines the generic data hiding process. Section "Watermarking Techniques" introduces and discusses the research works related to digital watermarking that exploit the properties of CI approaches. Furthermore, it analyzes the performance of the reviewed schemes by identifying their advantages and limitations. Finally, conclusions and future research directions are discussed in last section.

Background

To better clarify the context, in this section we provide broad definitions and discussions of the main topics of this paper including computational intelligence and digital watermarking.

Computational Intelligence

Computational Intelligence (CI) is a new area of research that studies bio-inspired adaptive mechanisms to enable or facilitate the creation of systems which exhibit intelligent behavior in complex and changing environments. Systems equipped with computational intelligence possess interesting characteristics such as adaptability, fault tolerance, and the ability to find in a reasonable time, solutions to hard problems with fairly low error rates [22, 23].

The research area spanned by the CI is wide, encompassing heterogeneous technologies such as fuzzy and rough sets, neurocomputing, evolutionary computing (EC), and chaos theory (CT). Nevertheless, in this paper we use the term CI in the most general sense, incorporating the fuzzy, neural and evolutionary computing methodologies, since these are the core of all CI's definitions. The CI technologies together with their related natural models are shown in Table 15.2. These approaches, except for fuzzy sets and chaos theory, are able autonomously to acquire and integrate knowledge, and can be used in either supervised or unsupervised learning mode.

Soft computing (SC) is another technology that has been recently incorporated under the umbrella of the CI. According to the general opinion, SC is "partnership" [23] of "hybrids or synergistic combinations/ensembles of these complementary approaches. Moreover, the resulting technique(s) will usually be iterative in nature, with successive solutions delivering improved performance/accuracy, to a user-specified degree." [24]. Thus, soft computing comprises hybrid intelligent systems such as, neuro-fuzzy, neuro-genetic, and fuzzy-genetic systems.

For a complete and up-to-date overview of the CI, together with its relations with Artificial Intelligence (AI), Machine Learning (ML) and Soft Computing (SC) fields, the reader can consult Fulcher's handbook [25]. Furthermore, for the application of the CI to multimedia processing and intrusion detection the reader is invited to look at the papers of Hassanien et al. [26] and Wu et al. [27].

Digital Watermarking

Nowadays, digital watermarking has emerged as the main technique to solve the fundamental issues of digital data, i.e., proof of ownership and content authentication. In general, watermarking techniques follow a generic model, which consists of two building blocks: the *embedding process* and the *extraction process* (also

Table 15.2 The CI technologies addressed in this paper and their related natural models

Computational intelligence technologies	Related natural model
Fuzzy sets (FSs)	Human reasoning processes
Artificial neural networks (ANNs)	Biological neural networks

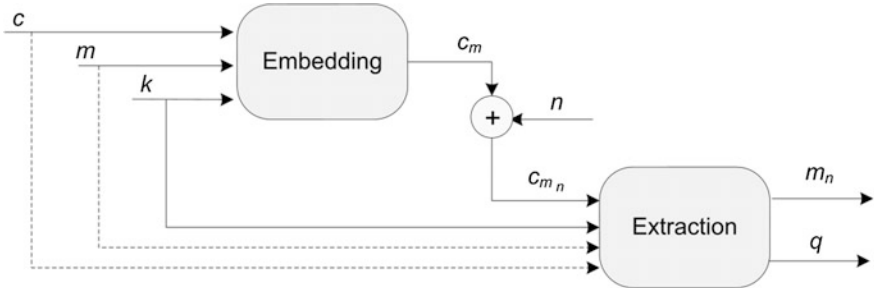


Fig. 15.1 Illustration of the generic watermarking hiding scheme: c represents the digital content, m denotes the secret message and k is the secret key

called detection process), as shown in Fig. 15.1. Each of them has a specific input, output and functionality in the overall scheme.

Three inputs characterize the *embedding process*: the digital content c, the secret message m to be hidden and a secret key k. The media content together with the secret message can be of any kind, e.g., pseudo-random sequences, text, images, videos or 3D models. The key k assures the security of the scheme and stores critical information about the embedding and extraction processes, such as the watermark generation process, the embedding strategy, the embedding locations, the quantization step etc. The output of the embedding process is the modified content c_m.

During the transmission, c_m can be corrupted due to various perturbations, the most common being noise addition or compression. The *extraction process* takes as inputs the noisy content c_{m_n}, the same secret key used during embedding and, depending on the data hiding scheme, the secret message together with the original content. In general, c_{m_n} and k are the principal inputs of the detector, whereas m and c are seen nearly as additional information, which in many cases increase the robustness of the entire scheme. The extractor can provide two outputs: first, the recovered message m_n, which, to assess its similarity, is further compared with m. The second output is a statistical measure q representing the probability that m is present in the noisy content. In order to preserve the imperceptibility of m and the quality of the watermarked media, the embedding energy of m is controlled by a strength factor.

The media content or the secret message is not always in a suitable representation; therefore, it is necessary to apply different operations to make it compatible with further manipulations. The main preprocessing operations are:

- Specific operations which are meant to modify c and/or m into sequences of random bits driven by special cryptographic tools, e.g., secure-hash functions (SHA), pseudo-random permutation functions, chaotic mixing, neural networks or Arnold transformation.
- Operations which split the media content in perceptual bands which are further used to devise masking models that reduce the visual impact of m. Most of these

models are based on the human visual and auditory system and their final goal is to maintain the secret message below a specified visual threshold. These operations are performed once before embedding, being unnecessary during the extraction.

- Endowment of the secret message with error correcting codes (ECCs) before the embedding process aims to improve its detection and robustness against attacks.

In the case of steganography, c represents the innocuous cover object used to share the secret message m through the stego-object c_m and k is called the stego-key. Instead, for digital watermarking m represents the watermark while the watermarked object c_m is the object that contains it.

Watermarking Techniques

The aim of this section is to provide a detailed overview of the watermarking algorithms that are based on several computational intelligence paradigms such as artificial neural networks and fuzzy sets. For the most recent research, there are currently several conferences dedicated to this topic such as the IEEE Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP). For a searchable library, we suggest the digital library of Applied Soft Computing Journal and Soft Computing Journal. In addition, the general conferences such as International Conference on Image Processing (ICIP) could have tracks related to intelligent information hiding.

Techniques Based on Artificial Neural Networks

An artificial neural network (ANN) is a processing system which mimics the behavior of the biological nervous system by means of a mathematical model which connects many neurons in a network fashion. Each neuron in a neural network (also called perceptron) performs a weighted sum of its inputs, e.g., x_1 and x_2, then it computes the output y_1 by applying a transfer function (TF) to the weighted sum. The output signal is further influenced by a threshold value θ, also referred to as the bias. The internal structure of a perceptron is depicted in Fig. 15.2. ANNs inherit the computational power and the cognitive abilities that biological neural networks possess.

A typically ANN comprises three layers: the input layer, the hidden layer and the output layer (see Fig. 15.3). The purpose of the hidden layer is to enhance the separation capacity of the network.

Between the neurons of each layer there are weighted connections which aim to propagate the signals or information through the network. The crucial issue of an ANN is how to set the values of the weights and the threshold θ such that a set of

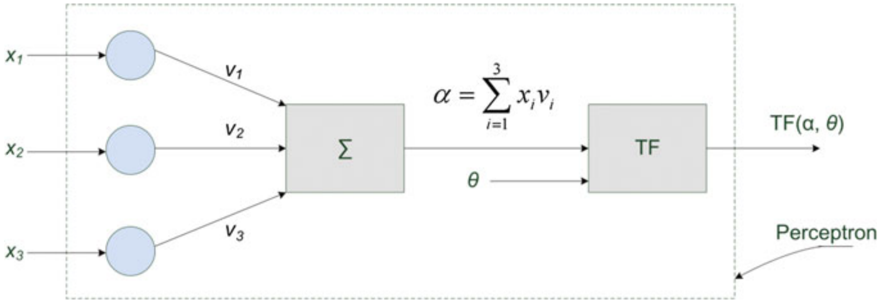


Fig. 15.2 Perceptron visualization: x_1, x_2, and x_3 are the inputs, v_1, v_2, and v_3 are the link weights, $\sum$ is a functional block which performs the weighted sum α, and TF represents the transfer function of the perceptron that receives α and θ, and determines the output of the neuron. Usually, TF is a linear or a sigmoid function

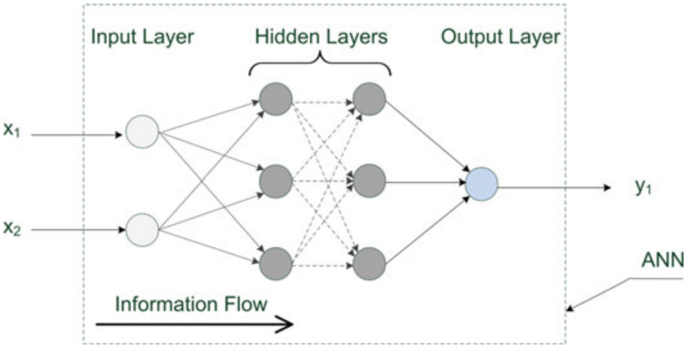


Fig. 15.3 Illustration of a 2-3-3-1 feed-forward ANN and the way the information travels across the network. The neural network has the following structure: one input layer with two neurons, two hidden layers each having three neurons (usually called hidden neurons), and one output layer with one neuron; x_1 and x_2 are the inputs while y_1 is the physical output of the network

desired outputs are obtained. This task is achieved by employing various learning approaches, which can be split into three classes: supervised learning, unsupervised learning and reinforcement learning. In supervised learning the NN is provided with a data set (i.e., training set) which consists of the input data and their target (desired) values. The purpose of this approach is to adjust the weight values such that the global error between the network outputs and the target values is minimized. Instead, in unsupervised learning there is no knowledge of the categories into which the inputs must be classified. Finally, reinforcement learning works by rewarding the efficient neurons while correcting the inefficient ones.

For each learning approach, several learning rules, e.g., gradient descendent rule, generalized delta rule and error-correction rule, have been developed. For a detailed description of ANNs, including their architectures, learning modes and transfer functions, the reader may consult [28, 29].

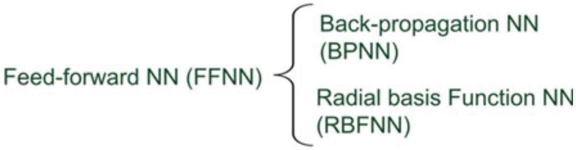


Fig. 15.4 The main ANN architectures reviewed in this paper

ANNs have been successfully applied in many data intensive applications including pattern recognition, function approximation, data mining, optimization and classification. In the context of data hiding, the properties of ANNs, such as learning ability and one-way behavior (i.e., the output of an ANN can be easily obtained by mixing the inputs of the neural network while being difficult to reverse this mechanism), have been extensively exploited.

The forthcoming section will review the ANNs contributions in devising watermarking schemes, and is organized depending on the network topology as illustrated in Fig. 15.4.

Feed-Forward Neural Networks

In the last decades, a lot of multi-layered neural network architectures have been devised. A well-known one is the feed-forward neural network (FFNN) in which neurons in each layer are fully connected merely with all the neurons in the next layer, causing information to flow in only one direction; thus, there are no feedback links within the network. Feed-forward networks can be broadly split into two classes:

- *Back-Propagation Neural Networks (BPNNs)*, which use the back-propagation learning techniques. Each learning iteration (epoch) consists of two steps: a feed-forward step that computes the output of the neural network and a backward propagation step which transmits the error from the output layer to the inner neurons.
- *Radial Basis Function Neural Networks (RBFNNs)*. Such networks were devised to work in both supervised and unsupervised learning modes. In particular, they use a nonlinear TF for the neurons belonging to the output layer while a linear TF for the hidden neurons.

Back-Propagation Neural Networks

The early application of neural networks to data hiding dates from the year 1999 when Yu et al. [30] in a short communication of the Signal Processing Journal proposed a neural network approach to improve the performance of the Kutter's watermarking scheme [31] for color images. To better understand how this improvement works, we decide to provide in Algorithm 1 a description of the embedding and detection procedures introduced in [31].

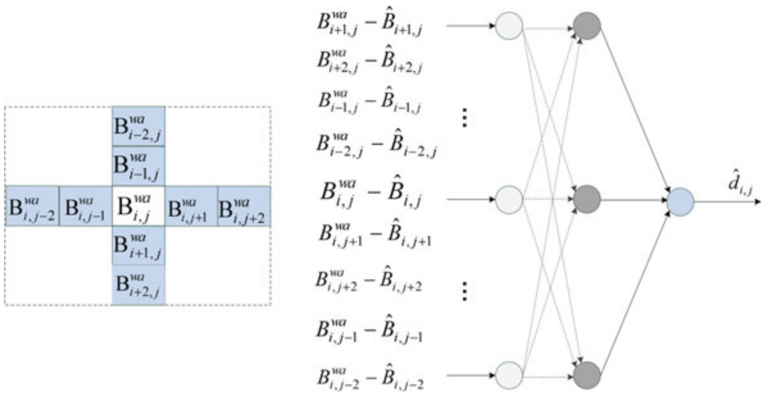
In essence, each watermark bit is inserted by modifying the blue component of a random pixel and then it is retrieved by taking the sign of the difference between the estimated value and the actual value of the pixel.

Algorithm 1: Kutter's Watermarking Algorithm

```
Input:  $I$, original color image; $H$, 2-bit sequence; $S$, digital signature of $I$; $k$, secret key.
Output: $I_w$, watermarked image.

1. Construct the watermark sequence of length $n$, i.e., $W=H+S$.
2. By means of $k$, generate $n$ pseudo-random positions over $I$.
3. To obtain the watermarked image $I_w$, for each pixel at the position $(i, j)$:
   3.1 Insert the watermark bit, $w_{i,j}$, by modifying the blue component of the pixel:

       $\begin{cases} B^{\mathrm{w}}_{i,j} = B_{i,j} + (2w_{i,j} - 1)L_{i,j}\beta \\ L_{i,j} = 0{,}299R + 0{,}587G + 0{,}114B \end{cases}$    (1)

       where $R$, $G$ and $B$ are the red, green and blue channels of $I$, $L_{i,j}$ is the luminance of the pixel, $B_{i,j}$ and $B^{w}_{i,j}$ is the blue channel of
       the original and watermarked pixel, and $\beta$ is the embedding strength.

Input:  $I_{wa}$, distorted (attacked) watermarked image; $H$, $k$.
Output: $W_r$, recovered watermark sequence.

1. Using $k$, generate the same pseudo-random locations within $I_{wa}$.
2. For each pixel at the position $(i, j)$:
   2.1 Compute an estimation of the original value of the pixel containing the watermark bit:

       $\hat{B}_{i,j} = \frac{1}{8}\left( \sum_{k=-2}^{2} B^{wa}_{i+k,j} + \sum_{k=-2}^{2} B^{wa}_{i,j+k} - 2B^{wa}_{i,j} \right)$    (2)

       where the $B^{wa}_{i+k,j}$ $(B^{wa}_{i,j+k})$ represent the blue channel of the pixels of $I_{wa}$, located in a cross-shaped neighborhood around the
       watermarked pixel at the position $(i, j)$.

   2.2 To recover the embedded bit, the sign of the difference between the estimated value and the actual value of the pixel, is
       considered:

       $w_{r_{i,j}} = \begin{cases} 1 & \text{if } B^{wa}_{i,j} - \hat{B}_{i,j} > 0 \\ 0 & \text{otherwise} \end{cases}$    (3)

       where $w_{r_{i,j}}$ is the recovered bit.
```

This estimation is built on a linear combination of the pixel values in a cross-shaped neighborhood around the watermarked pixel (see Fig. 15.5a). The Kutter's algorithm is attractive since it is blind, and has fairly good performances both in terms of quality of the watermarked image and robustness against common and geometrical attacks.

However, to increase the robustness and adaptability of the extraction process, Yu et al. [30] utilizes a neural network detector which consists of a 9-5-1 BPNN. Aiming at the correction of watermark bit extraction, firstly the NN needs to be trained. According to their method, the training set of the NN is composed of:


$B^{wa}_{i-2,j}$
$B^{wa}_{i-1,j}$
$B^{wa}_{i,j-2}$
$B^{wa}_{i,j-1}$
$B^{wa}_{i,j}$
$B^{wa}_{i,j+1}$
$B^{wa}_{i,j+2}$
$B^{wa}_{i+1,j}$
$B^{wa}_{i+2,j}$
$B^{wa}_{i+1,j} - \hat{B}_{i+1,j}$
$B^{wa}_{i+2,j} - \hat{B}_{i+2,j}$
$B^{wa}_{i-1,j} - \hat{B}_{i-1,j}$
$B^{wa}_{i-2,j} - \hat{B}_{i-2,j}$
$B^{wa}_{i,j} - \hat{B}_{i,j}$
$B^{wa}_{i,j+1} - \hat{B}_{i,j+1}$
$B^{wa}_{i,j+2} - \hat{B}_{i,j+2}$
$B^{wa}_{i,j-1} - \hat{B}_{i,j-1}$
$B^{wa}_{i,j-2} - \hat{B}_{i,j-2}$
$\hat{d}_{i,j}$

(a) Cross-shaped window

(b) ANN

Fig. 15.5 The cross-shaped window and the architecture of the neural network applied in [30]. The inputs of the NN are the estimated and real values of pixels within the cross-shaped window while $\hat{d}_{i,j}$ is the output of the neural network. Within the cross-shaped window, the central element represents the blue channel of the watermarked pixel

- An input vector with the differences between the real and estimated value of the watermarked pixel, i.e., $B^{wa}_{i,j} - \widehat{B}_{i,j}$, and its cross-shaped neighbors, i.e., $B^{wa}_{i+k,j} - \widehat{B}_{i+k,j}$ and $B^{wa}_{i,j+k} - \widehat{B}_{i,j+k}$ with $k = \{-2, -1, 1, 2\}$.
- A target vector that contains the normalized estimations of the original values of the pixels conveying the bits of H, and is defined by:

$$d_{i,j} = \frac{1}{255}(2h_{i,j} - 1)\widehat{B}_{i,j} \tag{4}$$

where $h_{i,j}$ are the bits of H. Note that unlike to Kutter's algorithm, in [] H is an m-bit sequence considered known in detection, since it is used in the training process. The structure of the neural network is shown in Fig. 15.5b.

After performing the training, a set of weights is obtained which is further used to recover the bits of S during the test phase. It is worthwhile to point out that, considering known a part of the watermark both the training and test phases of the NN are performed during the extraction. Experimental results carried over several test images show that the proposed scheme is robust to several attacks such as JPEG compression, low pass filtering, scaling and rotation. However, by using an estimation value ($\widehat{B}_{i,j}$) as the desired output (see Eq. (4)), the physical output of the neural network, $\hat{d}_{i,j}$, makes an estimation of $\widehat{B}_{i,j}$, instead of the original value, i.e., $B_{i,j}$, decreasing the accuracy of the network. Furthermore, due to the estimation performed by the NN it is not possible to completely retrieve the watermark sequence in case of no attacks. Finally, the details of the neural network and of the training process (e.g., the transfer functions of the hidden and output layers, the stopping criterion and the number of learning iterations) are not provided.

A similar neural network approach was used by Lu et al. [32] to devise a color watermarking scheme, which inserts the watermark by taking into consideration the correlation between the watermarked pixel and its square-shaped neighbors. Thus, the embedding rule becomes:

$$B^{w}_{i,j} = \overline{B}_{i,j} + (2w_{i,j} - 1)L_{i,j}\beta \tag{5}$$

where $\overline{B}_{i,j}$ is the blue channel average value over the square-shaped window centered at (i, j):

$$\overline{B}_{i,j} = \frac{1}{8}\left(\sum_{k=-1}^{1}\sum_{k=-1}^{1} B_{i+k,j+k} - 2B_{i,j}\right). \tag{6}$$

The authors state that this modification aims to improve the robustness since the original values of the watermarked pixels are better estimated by its square-shaped neighbors. Nevertheless, the experimental results show that the watermarking scheme exhibits similar performance with those provided in [30, 31], against common attacks.

In [33] a blind watermarking technique in spatial domain based on neural network is proposed. To enhance the security of the scheme, the watermark, prior to embedding, is encrypted by using a chaotic sequence. In the embedding process, the following operations are performed: (1) the original image is split into 3 × 3 nonoverlapping blocks; (2) for each block an 8-10-1 BPNN is applied to learn the relationships among the pixels; (3) the encrypted watermark is inserted into the central pixel $p_{i,j}$ of each image block according to the equation:

$$p^{w}_{i,j} = \begin{cases} p_{i,j} - \beta & \text{if } \hat{p}_{i,j} - p_{i,j} \le \beta \text{ and } w_{i,j} = 0 \\ p_{i,j} + \beta & \text{if } \hat{p}_{i,j} - p_{i,j} > \beta \text{ and } w_{i,j} = 1 \\ \min(p_{i,j}, \hat{p}_{i,j} - \beta) & \text{if } \hat{p}_{i,j} - p_{i,j} > \beta \text{ and } w_{i,j} = 0 \\ \max(p_{i,j}, \hat{p}_{i,j} + \beta) & \text{if } \hat{p}_{i,j} - p_{i,j} \le \beta \text{ and } w_{i,j} = 1 \end{cases} \tag{7}$$

where $p^{w}_{i,j}$ denotes the watermarked pixel value, $\hat{p}_{i,j}$ is the output of the NN normalized in the range [0, 255] and β is the adaptive strength factor, which is computed for each block based on the signal-noise-ratio (SNR) between the original and watermarked image. The visualization of the scheme given in [33] is depicted in Fig. 15.6.

Since the neurons of the network use sigmoid transfer functions, the pixels of each image block are normalized in the interval [0, 1] dividing by the maximal

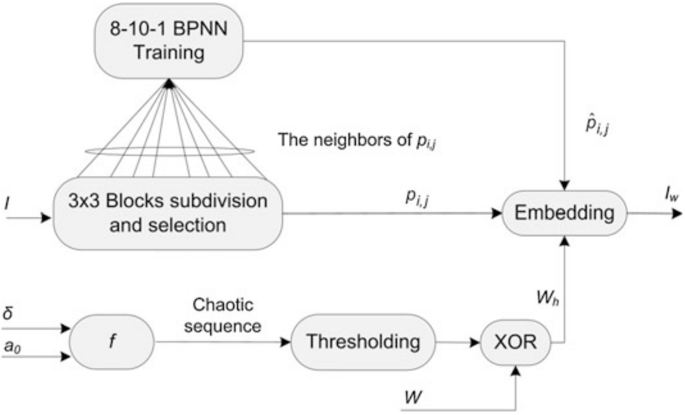


Fig. 15.6 The embedding scheme reported in Tang and Liao [33]. The inputs of BPNN are the neighbors of $p_{i,j}$ while the desired output is $p_{i,j}$ (i.e., the central pixel of the block). The chaotic watermark W_h is generated by XORing W with a thresholded chaotic sequence obtained from the skew tent map f with parameter δ and initial value a_0

amplitude, before being processed by the BPNN. In order to recover the watermark, the possible marked pixels are compared with the output of the trained BPNN. Therefore, the method is capable to extract the watermark without the original image. As pointed out by the authors, an important aspect of the watermarking scheme is the segmentation size which directly affects the input of the NN and the capacity of the watermark while it is tested when the watermark is recovered. The experimental tests carried out show that the similarity ratio between the recovered and original watermark is not less than 85% if the watermarked image undergoes common signal processing operations, such as JPEG compression, filtering, histogram equalization, noise addition and cropping.

Another difference from the approach used in [30] is related to the NN, that is trained during embedding. In addition, the quality of the watermarked image reported is very high, i.e., PSNR = 48.55 dB. The potential problem with this method is the impossibility to extract the watermark bits without errors, since the recovered watermark can only be obtained approximately [33].

Inspired by Yu et al. [30], Zhang and Wang [34], Sang and Alam [35] presented a lossless image watermarking scheme in spatial domain based on neural network. Actually, this scheme does not embed any information into the original image, *I*. It rather constructs a secret key as the bitwise exclusive-OR of the watermark (in this case an authentication code) and a binary sequence derived from the features of *I* with the aid of a neural network. In order to generate the binary sequence, firstly a number of pixels are randomly selected from *I*. Then each selected pixel is approximated by its square-shaped neighbors using an 8-10-1 BPNN. Finally, the binary sequence is obtained by comparing the output of the neural network (i.e., the estimated value of the selected pixel) with its corresponding desired output (i.e., the real value of the pixel).

To recover the watermark from a test image, the secret key is XORed with a binary sequence obtained by employing the above procedure. Since there is no watermark embedded into the original image the scheme cannot be considered lossless, i.e., "after the embedded information is extracted, we can revert to the exact copy of the original image before the embedding occurred [36]," and neither may be used for copyright protection. The main downsides of the scheme are "slow computation and the extra storage needed [35]." The authors also published an extended version of this paper [37] based on the same approach.

Exploiting the properties of the neural network, Lian [38] proposed a novel image authentication technique. To achieve this goal, first the image is segmented into nonoverlapping blocks; then, for each block, a single layer NN is constructed and fed with the secret key, the blocks' pixels, and the watermark, which in this scenario acts as an authentication code for the image (see Fig. 15.7). By mixing these quantities, the neural network produces a secret parameter which can be later used to authenticate the image.

The interesting thing with this scheme is related to how the parameters of the neural network, such as the weights, the inputs, the secret parameter, the output and the desired value, are set in order to authenticate the image. More precisely, these parameters are defined as follows:

- The weights of the NN are generated by a PRSG with an initial state which depends on the value of the secret key.
- The output and the desired value of the network are the same, and are given by the watermark. Since the output and the weights of the network are already known, this NN will not need to be trained or tested.
- The elements of the input vector are the pixels of the image block.
- The secret parameter b is an unknown value which is computed by combining the other features of the neural network, i.e., the input vector, the weights and the output value.

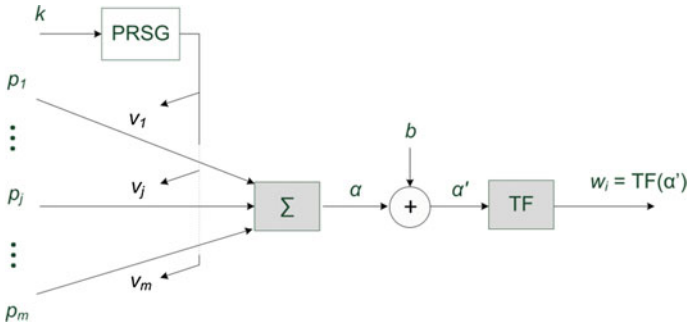


Fig. 15.7 The perceptron used by Lian [38] to authenticate each block of the image. The meaning of the symbols is: p_1, …, p_m represent the pixels of the block, v_1, …, v_m are the weights of the NN obtained through a pseudorandom sequence generator (PRSG) which uses the secret key k as the random seed, b denotes the secret parameter and w_i is the i-th bit of the watermark

The proposed scheme is robust to signal processing operations (noise addition and compression) and to malicious content modifications. By analyzing the relation between the block size and the performance of the scheme, the author gives several indications to improve the detection of the authentication code. Finally, the algorithm is fast since it works in the spatial domain and doesn't train nor test the neural network.

Piao et al. [39] split the original image into nonoverlapping blocks and then each block undergoes two different operations. The first operation is performed into spatial domain and it computes the average pixel value of the block while the second operation applies the DCT transform on the block. In order to embed the watermark data into each block a trained neural network is used with quantized and rounded value of the DC component as the input and the average pixel value of the block as the desired output (see Fig. 15.8).

In the paper, several implementation details of the BPNN are offered, including the activation functions for the hidden and the output layers, the learning rule, the number of epochs and the training error. The purpose of the neural network is to estimate the average value of the pixels within each image block, and then to use this estimation in the embedding and extraction rules. Thus, by incorporating the output of the NN the embedding rule becomes:

$$p_{i,j}^{w} = \begin{cases} p_{i,j} + round\left(\dfrac{2Q + 64\widehat{M} - M}{M}\right) & \text{if } w_{i,j} = 1 \\ p_{i,j} + round\left(\dfrac{-2Q + 64\widehat{M} - M}{M}\right) & \text{if } w_{i,j} = 0 \end{cases} \tag{8}$$

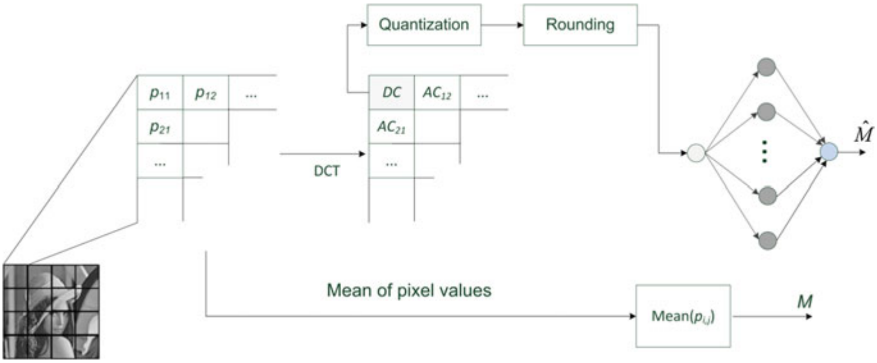


Fig. 15.8 The training process of the BPNN used in [39]: $p_{i,j}$ are the pixels of an 8 × 8 block, $\widehat{M}$ is the physical output of NN and M is the desired output for the NN which represents the average pixel value of the block

where $w_{i,j}$ is the watermark bit. In the extraction, the watermark is recovered by comparing the mean of each block of the test image with the output of the trained NN. Experimental results show that the method has better performances compared to the schemes proposed by Chen and Wornell [40] and Mei et al. [41] in terms of imperceptibility, robustness against JPEG compression, noise and resizing.

All the algorithms discussed earlier embed the watermark by modifying the luminance values of the image pixels. However, this approach is not suitable for devising robust watermarking since in most cases spatial features can be easily modified by signal processing operations. Being aware of this limitation, several researchers shift attention to transform domain watermarking. For instance, by using a neural detector Tsai et al. [42] proposed a blind version of the Swanson's audio watermarking scheme [43] which is based on the properties of the human auditory system and DCT. In [43] the use of the audio content during extraction is compulsory and is due to the adopted embedding rule:

$$t_{i,j}^{w} = t_{i,j} + \beta w_{i,j} \tag{9}$$

where $t_{i,j}$ is the original DCT coefficient, $w_{i,j}$ the watermark bit, β is the strength factor and $t_{i,j}^{w}$ is the watermarked DCT coefficient. Hence, to extract the watermark the DCT coefficients of the audio signal are needed. To overcome this limitation Tsai et al. [42] use a 9-9-1 BPNN to learn the original values of the DCT coefficients that are marked. The architecture and the training process of the NN are similar to that proposed in [30]. During the extraction, the trained network is put to work to estimate the original DCT coefficients. To extract one bit of watermark the sign of the difference between the estimation and the real value of the DCT coefficients is considered. The experimental results performed over three audio signals demonstrate the robustness of the scheme against multiple signal processing operations. As we noted for other techniques, the Tsai's scheme [42] is not able to recover the watermark bits without error even if the watermarked image has not suffered any modifications.

Lu et al. [44] adapts the DCT-based watermarking approach introduced in [45] by incorporating a neural network decoder to recover the watermark. During embedding process, the original image is subsampled in four subimages followed by the application of DCT on each of them. The watermark data is embedded into a set of DCT coefficients according to the following equation:

$$t_{i,j}^{w} = t_{i,j} + t_{i,j}\left(2w_{i,j} - 1\right)\beta \tag{10}$$

which is similar to that proposed by Kutter et al. [31], the only difference consisting in the use of the DCT coefficients instead of the pixels values. During the tests the watermarked images, i.e., "Lena" and "Baboon", are subject to single and combined processing operations including JPEG compression, filtering, noise addition and their related combinations. The proposed scheme can extract with high accuracy the watermark from all the attacked images. Two issues related to the main requirements of the watermarking scheme raised our attention. Firstly, the PSNR values of both watermarked images are not specified. Secondly, there are quite notable differences between the performances of the scheme obtained for the "Lena" image and those for the "Baboon" image. Hence, from the reported results is difficult to draw any conclusion regarding the performance of the scheme.

In general, the characteristics of a watermarking scheme are closely related to the application the scheme is intended to serve. There are applications for which the set of admissible attacks can be easily anticipated. For example, in case of sharing medical images for e-health related applications, the main concern is the JPEG compression, used to reduce their storage. Similarly, noise addition is another highly probable attack which affects many applications. Driven by these considerations, Khan et al. [46] present a novel watermarking scheme which intelligently exploits the learning abilities of support vector machines (SVMs) and ANNs to devise decoding structures that take into consideration the intended application and the expected attack. The embedding framework is that introduced by Hernandez et al. [13] which employs DCT in blocks of 8 × 8 pixels and inserts the watermark additively into the middle frequency coefficients, which can be approximated by zero-mean generalized Gaussian probability density function [47]. To generate the watermark, a hidden message m, i.e., a bipolar signal with values $\{-1, 1\}$, is firstly encoded into a codeword vector and further multiplied by a sequence generated using a pseudorandom generator seeded with the value of the secret key k. The perceptual mask β is obtained through a perceptual model which takes into consideration the frequency masking properties of the human visual system (HVS). The purpose of this mask is to control the energy of the watermark during its insertion. The block diagram of the embedding process used in [13] is given in Fig. 15.9.

The message is decoded by employing a maximum likelihood based decoder structure which calculates the sufficient statistics of each inserted bit and then compares it with a threshold. The main assumption is that the probability density function of the DCT coefficients remains unchanged even after an attack is applied on the watermarked image. However, this detection approach is incapable to extract the message bits efficiently since the distributions of sufficient statistics overlap considerably in case of an attack. To resolve this issue, Khan et al. [46] suppose that

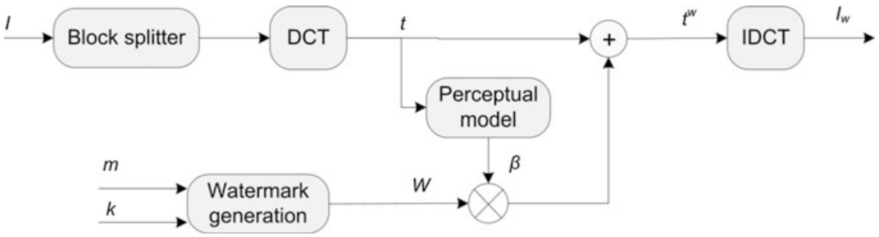


Fig. 15.9 The embedding scheme proposed by Hernandez et al. [13]: t and t^w are the original and watermarked DCT coefficients, respectively

"a non-separable message in lower dimensional space might be separable if it is mapped to a higher dimensional space," attained by the hidden layers of the ANN and the kernel functions of the SVM. As consequence, the decoding is regarded as a classification problem where a bit relates to a sample and a message corresponds to a pattern. To extract the features corresponding to each bit of the message two methods are implemented. The first one, adopted by Hernandez et al. [13], computes the sufficient statistics across all frequency bands. Instead, Khan et al. [46] calculate it across the channel. By considering the frequency as the feature, each message pixel can be characterized by 22 features, i.e., 22 middle frequency bands selected to convey the message. Hence, each sample in the training set comprises: the input of 22 features and the related desired value (i.e., the original embedded bit). The parameters of the ANN are: the input layer of 22 neurons, two hidden layer of 25 and 15 neurons, and an output layer of one neuron. To assess the performance of the SVM, ANN and Hernandez et al. [13] detectors against Gaussian noise, JPEG compression and Winner attack, data sampling techniques, such as self-consistency and cross-validation, are employed. Contrary to the SVM and the maximum likelihood decoders, the ANN detector has worse performance on the test data. This is due to the low generalization capability of the ANN model. Furthermore, the performance of ANN and maximum likelihood decoders depends greatly on the characteristics of the original image. Apart from this, the temporal cost induced by the proposed decoding structures lower their applicability, e.g., in case of Gaussian attack, the training time of the ANN decoder is 30 min while SVM decoders need between 15 and 19 min to complete the learning process.

The watermark extraction approach introduced by Tsai et al. [42], i.e., the network inputs are the neighbors of the watermarked coefficients while the physical output is an estimation of the original value of the watermarked coefficient, was also extended to watermarking algorithms based on DWT. For example, Xu et al. [48] reported a novel scheme which employs the statistical properties of the DWT to invisibly embed the watermark data into the textured and edginess areas of the image. Instead, Li et al. [49] and Wang et al. [50] devised watermarking schemes which make use of the wavelet moment modulation to achieve resistance to geometric attacks. Both schemes exhibit good performance against a wide range of attacks, but the computational and temporal costs attained by these techniques could hardly be accepted for real-world applications.

Research, such as [51–54], exploited the learning capabilities of the neural networks for purposes different from watermark extraction. To improve the performance against attacks, Lu et al. [54] employ a BPNN to distinguish the characteristics of the extracted watermark even when it can be hardly recognizable. However, we are very interested in the false positive rate of the classification system since even if the normal correlation of the extracted watermark is 0.1628 the trained neural network still associates it with the original watermark.

In case of the fragile watermarking scheme by Fan et al. [51], the NN is used to identify what kind of distortion has been occurred.

Specifically, by using the features of the difference between the original watermark and the extracted one a BPNN is capable to assess if the perturbed image was attacked by cropping, salt&pepper noise or scaling. However, the lack of details regarding the parameters of the algorithm makes very difficult its implementation and evaluation.

Instead, Zhao et al. [52, 53] devised two watermarking methods in the wavelet domain that use neural networks for scrambling (a) the chaotic sequence through which the watermark is encrypted [52] or (b) the watermark image prior to its embedding [53]. In the first scheme, the watermark is encrypted by combining the chaos theory and the neural network. Instead, the latter scheme scrambles the watermark as follows: it splits the watermark into nonoverlapping blocks and the pixels of each block are taken as the inputs of the neural network; next, the neural network is trained so that the output of the hidden layer scrambles the watermark while the outputs of the network provide the original watermark (i.e., it reverses the scramble).

In both schemes, the encrypted watermark is embedded into DWT coefficients of all frequencies. Although, the authors sustain that both methods are credible and robust against several attacks like, addition of Gaussian and salt and pepper noise, and JPEG compression, none of the papers provide any experimental results.

Realizing the importance of choosing the embedding space, Elbasi and Eskicioglu [55] employ a classification system based on neural network in order to select the optimal transform, e.g., DCT, DWT, and DFT, for watermark embedding. The experimental results show that the accuracy of the classification system is very promising.

Radial Basis Function Neural Networks

In a radial basis function (RBF) network each neuron of the hidden layer represents a different RBF, that is characterized by a number of parameters. This type of neural network performs classification by computing distances between inputs and the RBF centers of the hidden neurons. RBFNNs are much less used in the context of data hiding although they are much faster than the BPNNs [27]. Another interesting aspect is that in most cases the RBFNNs are coupled only with the watermarking schemes which embed the watermark in DCT- or DWT-domains [56–60].

In 2003 Zhang et al. [56] proposed a block-wise watermarking scheme based on the DCT and RBFNNs. The watermark, i.e., a meaningful binary image, is inserted adaptively into the middle frequency coefficients of the image blocks. By taking into consideration the frequency component, a RBFNN is used to find the optimal embedding energy of the watermark. It is important to mention that the adaptive strength factor provides a greater robustness against common image processing operations and a better quality for the watermarked image.

Lu et al. [57] decompose the host image into nonoverlapping blocks and apply DCT on each of them. For security purposes, the DCT coefficients in each block are pseudo-randomly permuted using a secret key. At this point, from each image block a series of low AC coefficients are selected with aim of training/testing the RBFNN

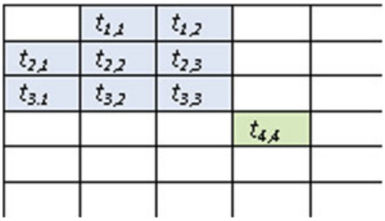


Fig. 15.10 The DCT coefficients of each image block of 8 × 8 pixels used by the 8-5-1 RBFNN: $t_{1,1}, \ldots, t_{4,4}$ denote the low frequency coefficients used to train and test the NN

and for watermark embedding. The set of the chosen coefficients is shown in Fig. 15.10.

In the training phase of the network, a number of image blocks, called the training blocks, together with their corresponding coefficients are used to construct the training samples: for the i-th block, $t^i_{1,1}, \ldots, t^i_{3,3}$ are the inputs of the RBFNN while the desired value is given by $t^i_{4,4}$. Once completed the training, the watermark bits are inserted into a particular coefficient of the remaining blocks, called test blocks, using the following relation:

$$t^{j^w}_{4,4} = \hat{t}^j_{4,4} + \beta \cdot w_j \tag{11}$$

where $\hat{t}^j_{4,4}$ is the output of the trained RBFNN for the j-th test block, $t^{j^w}_{4,4}$ is the watermarked coefficients and β is the strength factor. Thus, the purpose of the RBFNN in the embedding rule is to replace the real value of $t^j_{4,4}$ with its estimation, $\hat{t}^j_{4,4}$. By adopting this approach, the scheme attains good robustness against common and geometrical attacks.

The drawbacks of this method are:

- the necessity of training and testing the RBFNN in both embedding and extraction phases.
- there is a tight connection between the learning accuracy of the network and the watermark capacity. Specifically, in order to improve the learning rate is necessary to increase the number of training blocks which implicitly lowers the number of test blocks used to convey the watermark data.

Influenced by its earlier paper [39], Piao et al. [58] attempt to use a neural network to learn the characteristics of the DWT coefficients. To achieve this, firstly the original image is decomposed by a 4-level DWT transform, and then each coefficient of the 4-th level is approximated by a RBFNN (see Fig. 15.11).

After completing the training, a set of coefficients from the 4-level are selected for embedding through a secret key. Next, each selected pixel is replaced by its approximation value followed by the additive insertion of the payload. The experimental test carried out might not be sufficient to proof the performance of the algorithm. Another issue is that there is no satisfactory motivation of the training set

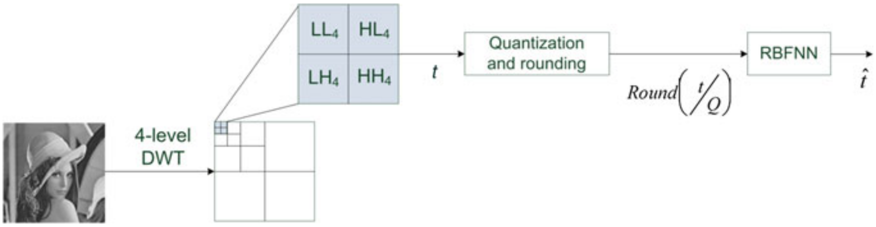


Fig. 15.11 The training process introduced in [58]: t is the desired value of NN and represents the DWT coefficient of the LL_4, HL_4, LH_4, or HH_4 subbands, $\hat{t}$ is the output of the neural network and Q is the quantization step

selection. A similar neural network approach has been also employed in the Zhang's watermarking scheme [59].

In [60], authors introduce a novel video watermarking scheme based on 3D-DWT and RBFNN. Embedding is performed by adding the watermark bits to the LL sub-band wavelet coefficients of the video frames. In order to maintain the quality of the video content, the watermark strength factor is adaptively computed. The neural network is employed in both embedding and extraction, and aims to memorize the relationships between watermarked wavelet coefficients and their neighbor's, as in [48, 49]. Experimental results show that the embedded watermark is robust and invisible. The main disadvantage which can be invoked is its highly temporal cost introduced by the 3D-DWT and the neural network.

Considerations

In this section, we reviewed research contributions on ANNs for digital watermarking. These research works take into consideration the neural networks' capability to generalize from limited, perturbed, and incomplete data, i.e., the digital content. Several ANN architectures were employed for resolving different watermarking tasks. Among them, the extraction of the watermark from noisy digital contents became the most significant aim for artificial neural networks [30, 32–35, 37–39, 46]. By adopting the NN in the extraction phase, the schemes gain several highly desirable properties, such as blindness and adaptability to attacks. Thus, instead of resorting to the original content in the extraction, researchers make use of a NN along with a set of weights to estimate the original content or even to directly extract the watermark.

Even if there are numerous watermarking scheme based on this idea the majority of them construct the training set of the NNs in the following manner:

- The inputs of the NN are the neighbors of the watermarked coefficient (pixel) while the target value is the value of the original coefficient (pixel) [33, 48, 57]. Another possible variation is to consider also the watermarked coefficient among the inputs of the NN [42].

- A second major possibility is given by the work of Tsai et al. [30] which considers the desire value of the NN as the original bit of the watermark. In this case, each input of the NN denotes the difference between pixel's (or coefficient's) value and its neighbors [33, 44, 50].
- Unlike to previous approaches, in Khan et al. [46] each sample of the training set represents 22 features and the corresponding target value (the original embedded bit). Thus, the extraction of the watermark bits becomes a binary classification problem.

Moreover, the domain of the features that constitute the input of the NN is another aspect which differentiates the research work based on ANN. Besides watermark retrieval, ANNs have been employed to: (a) encrypt the watermark [52, 53]; (b) compute adaptively the watermark strength factor [41, 56]; (c) investigate how a media content has been corrupted [51]; (d) choose the optimal domain transform for watermark embedding [55]; (e) validate the meaningful extracted watermarks which are hardly recognizable [54].

The watermarking schemes based on ANNs, except that proposed in [46], reveal the following issues:

- *Lack of information regarding development of ANN models*. Unfortunately, most of the authors make use of the ANN models in their algorithms without providing sufficient implementation details, such as the activation function of the hidden and output neurons, the training algorithm, the number of epochs, and the stopping criterion, leaving the reader to guess them.
- *Validation of the ANN models.* The NN models are not assessed by employing both cross-validation and self-consistency data sampling techniques. The purpose of applying self-consistency test is to verify the performances of ANN models on the training set while cross-validation test is used to evaluate the models on the novel data samples, i.e., on the training set. Without using these data sampling techniques, phenomena which lead to poor predictive performances, like overfitting and underfitting, cannot be avoided. For example, we found a particular flaw which affects all the schemes [32, 44, 48–50] which are based on the Yu's scheme [30]. More precisely, *both* the training and the test phases are performed during detection. The flaw is caused by considering, *within the same image*, several pixels (or coefficients, if the extraction is carried out in the transform domain) for training and others for testing. It is obvious that the NN is trained and tested on the image used in detection which may be also a perturbed image or even a fake image. Furthermore, in case of feeding the detector with attacked or fake images, by increasing the training samples the model does not increase its learning abilities.
- *Performance analysis*. The experimental tests performed by the research works are incomplete to draw any conclusion. Among the most important missing information we can mention: the quality of the watermarked image, i.e., PSNR value, the structure of the test set, the training time, the bit error rate (BER) ratio and the parameters of the attacks. Moreover, since all techniques are

block-based there is no analysis of the relation between the segmentation size and the learning capabilities or the watermark capacity.

- *Comparative analysis.* The majority of the works aim to improve some watermarking algorithms [31, 34, 40, 41, 43] which do not have strong theoretical foundations, as that proposed in [13]. Therefore, the comparative analyses performed are not enough to demonstrate that the "improved" schemes (based on ANN) have better performance than their counterpart schemes. This fact is confirmed also by Khan et al. [46] which show that, on the test data, the performances of the ANN detector are worse than that of the SVM model and the correlation-based model.
- *Detection.* As we mentioned before, the main purpose of using the ANNs is to eliminate the need of the original content during the extraction process. However, even so the watermarking schemes are not blind, as wrongly the authors sustain. At most these schemes are semi-blind which implies the utilization of the watermark or other side information (derived or not from the original content) [3, 61, 62].
- *Applicability.* In order to retrieve the watermark, the trained NN need to be provided during detection. Depending on the intended applications, e.g., copyright verification, content authentication and device control, the trained model can be made public or transmitted through a secure channel. Thus, by constraining the access to the training model will increase the overall security of the watermarking system but, in the same time, will restrict its applicability [46].
- *Security analysis.* In case of a digital watermarking algorithm security is strongly related to protecting its secrets, i.e., the secret keys, the topology and the weights of the NN, the segmentation size, etc. Since playing with the trained NN is permitted during extraction, a malevolent user may disclose the NN weights by feeding each neuron with *properly chosen inputs*, attack which is deeply analyzed in oblivious NN computing protocols [63–66]. Even if this issue is extremely important, only Lian [38] and Khan et al. [46] have proposed several solutions. In [46] the security of the watermarking scheme relies on the NN "inherent property of transforming the input vector to higher dimensional space being very difficult for an attacker to gain knowledge about the key" [46] by analyzing the output of the NN. Instead, Lian [38] imposes a lower bound on the size of the authentication image in order to avoid the disclosure of the NN weights by applying a brute-force attack.

Techniques Based on Fuzzy Sets

Fuzzy logic is a concept that was introduced by Zadeh [67], being appropriate to describe the vagueness in linguistics, and in addition capable to articulate human knowledge and inference ability in a transparent manner. To be able to deal with uncertainty and imprecision in input data, fuzzy logic uses an important concept namely, the fuzzy set. A fuzzy set is a set which do not have a clearly defined boundary, i.e., the contained elements have partial degrees of membership. The way the inputs points are mapped to their degrees of membership, which take values in range [0, 1], is given by the membership function. Instead, in a classic (crisp) set a collection of elements can belong or not to the set.

Fuzzy inference systems (FISs) are used to model the approximate way of reasoning which is a crucial characteristic of humans that give them the ability to make decisions in an inaccurate environment. To achieve this goal, the FIS applies a set of fuzzy rules (i.e., if-then rules) which forms the rule base of the system. The fuzzy image approach consists of a rule-based system which employs fuzzy logic to process the image data. The membership function can be defined for whole image or locally, and expresses a certain peculiarity of the image such as texture, luminosity and edginess. Thus, due to their properties, like the ability to handle uncertain and imprecise data, fuzzy systems have been employed to various image processing applications including image segmentation, image coding and noise reduction [68–71].

A good watermarking scheme has to exploit the particular properties of the image being watermarked in on order to embed imperceptibly the watermark. However, a watermark can have different levels of imperceptibility, leading to a greater or lesser likelihood for a given observer. Automated perceptual models, i.e., the human auditory system (HAS) [72] and the human visual system (HVS) [73, 74], which take into account various perceptual phenomena, such as sensitivity and masking, are used to measure and control the perceptibility of the watermark [3].

Watermarking schemes that endeavor to modify the watermark according to some perceptual models are referred to as perceptually adaptive schemes. Generally, the perceptual models are applied during the embedding process for automatic modification of the local or global embedding strength in order to achieve a specific perceptual distance between the original and watermarked image. We now discuss the most representative adaptive watermarking schemes which use the fuzzy sets to construct different perceptual models, also referred to as fuzzy perceptual masks (FPMs) [75, 76].

One of the earliest adaptive watermarking schemes based on a perceptual fuzzy model was introduced by Tao and Dickinson [77]. The block-DCT domain is used for watermarking due to its energy compaction capability and suitability to integrate the HVS features. Specifically, the watermark is inserted into several AC coefficients having the smallest quantization step, i.e., the frequencies of high energy. The essential idea is to assign a noise sensitive label to each image blocks of 8×8 pixels, and then to use it for watermark insertion. The classification algorithm [78] exploits the luminance, the edge and the texture masking effects. Each image block is classified into one of the six perceptual categories: edge, uniform with low intensity, uniform with moderate or high intensity, moderately busy, busy and very busy. However, since the test used to identify the edge blocks was too general (i.e., a block is an edge block only if its gradient magnitude is greater than that of the entire image) the authors proposed an additional test, called ratio test, which considers the variances of the neighboring blocks of the current block. Another interesting aspect of the classification procedure is the differentiation between the uniform blocks with moderate intensity and those with high or low intensity [77]. To extract the watermark blindly and without using the fuzzy perceptual mask, an optimal correlation-based detector is employed. However, there are some issues with this method: firstly, it does not report the PSNR values for the watermarked images and secondly, the reported results related to the robustness against JPEG compression and noise addition are not compared with other watermarking schemes.

The approach introduced in [77] has influenced several watermarking schemes such as those proposed by Mohanty et al. [79], and Kankanhalli and Ramakrishnan [80]. In these schemes, the FIS is used to compute two weighting factors, both applied in the embedding rule: the scaling factor, used to control the energy of the DCT coefficients, and the embedding factor, used to strength the watermark. An important inconvenience is that the original image is required to retrieve the watermark data. Furthermore, we could not figure out the reason of using a perceptual mask since these watermarking techniques insert *visible* watermarks into the host images.

Huang and Shi [81] proposed an adaptive spread-spectrum watermarking scheme based on DCT and HVS. By considering the brightness and the spatial sensitivity, the FIS classifies the image blocks into three classes, i.e., dark and weak texture class, bright and strong texture class, the class of the remaining blocks, and allocates to each of them adaptive strength factors. To convey the watermark, the low frequency coefficients of each block are selected. Aiming to improve the quality and robustness of [81], Lou et al. [82] refine the fuzzy inference system by classifying the blocks into six classes, to satisfy more smoothly and accurately the human visual model.

In [83, 84] the fuzzy perceptual model is developed in spatial domain and aims to exploit the spatial masking, the intensity resolution and the intensity sensitivity. The adaptive block processing technique, which is similar with that applied in [85], consists in:

- splitting the host image in blocks of 8×8 pixels;
- segmenting each block into five subblocks, as shown in Fig. 15.12;
- inserting the watermark by uniformly modifying the pixels of the central subblock B;

The embedding process of the DCT-domain watermarking scheme used in [83, 84] is illustrated in Fig. 15.13. It is worth to point that the pixel adjustment applied to embed the watermark is given by the FIS which take as inputs the mean intensity of the pixel belonging to each adjacent subblocks, i.e., the B_i with $i = 1, \ldots, 4$. Each input is composed of three membership functions (black, grey and white) while the output of the FIS, the pixel adjustment, appertains to minimum, medium and maximum membership functions. The applicability of the scheme is however

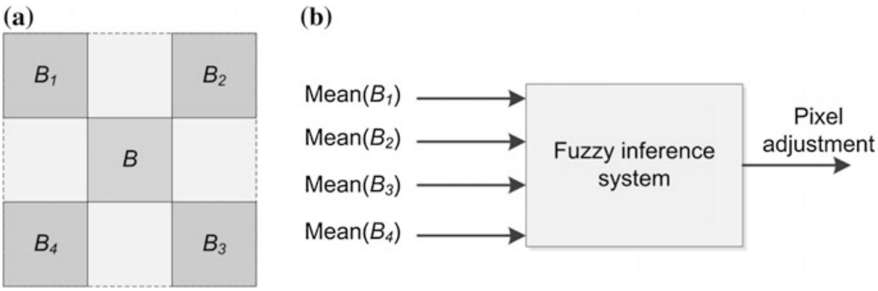


Fig. 15.12 The segmentation method used for each image block and the FIS employed in [83, 84]. The watermark is embedded into subblock *B* by adjusting its pixel values

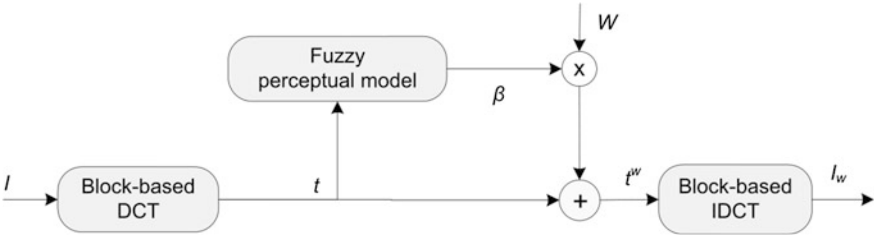


Fig. 15.13 The watermark embedding process with a fuzzy perceptual model used in [83, 84]

reduced since the original image is needed in the extraction process. Although, such scheme claims higher robustness against attacks, the comparative tests show that the results related to resampling and low pass filtering attacks are worse than those reported in the Lee's scheme [85].

In [75] an adaptive watermarking technique in wavelet domain which exploits the HVS and the FIS is developed. To compute a particular weighting factor for each coefficient of the DWT sub-bands, the FIS accounts for three phenomena: brightness, texture and edge sensitivity. The texture and the edge sensitivity are computed for a specified window which contains the wavelet coefficient. In the extraction, instead of approximating the perceptual shape and retrieve the watermark, the authors prefer to resort to original image. A similar fuzzy approach was used by the authors in a follow-up paper [76].

Hsieh and Tseng [86] apply a fuzzy inference filter to calculate the contextual energy of each coefficient in a DWT sub-band, and choose the coefficients with the largest contextual energy to carry the watermark. In order to calculate the contextual energy corresponding to each coefficient in a sub-band the neighbor coefficients are considered, as in [75, 76, 87]. The potential issue of the method is that it needs to store additional information (the value of original coefficients) to recover the watermark.

Considerations

In case of the digital watermarking, the fuzzy logic approach is employed to obtain different strength factors for the watermark by exploiting the local characteristics of the image such as frequency and brightness sensitivity. Moreover, fuzzy sets aid to improve the performance of some machine learning algorithms such as c-means clustering algorithm [88, 89].

The features of the fuzzy-based watermarking schemes can be summarized as follows:

- *Fuzzy rules*. The fuzzy rules of the inference system have been devised, mostly in the DCT domain [79–82], in the spatial domain [77, 78, 83–85] and in the DWT domain [75, 76, 86, 87]. In addition, these rules are generated manually instead of creating them automatically by applying for instance artificial neural networks [90, 91] or genetic algorithms.
- *Block-based processing*. All reviewed papers segment the host image into nonoverlapping bocks, and use the FIS to compute local strength factors for the watermark.
- *Detection*. Except for the work proposed by Tao and Dickinson [77] which is blind and does not invert the perceptual shaping in the detector, the other algorithms use the original image to reconstruct the fuzzy perceptual mask. Furthermore, the estimation of the fuzzy mask by using the received image (possible watermarked and noisy) instead of the original image in the FIS is not addressed by any scheme.

From the research work reviewed in this section we conclude that the utilization of fuzzy logic is still in infancy, deserving more attention in the near future.

Conclusions

The use of computational intelligence techniques in the watermarking context is an exciting research direction, which attracts considerable interest. Due to their characteristics, such as adaptability and fault tolerance, intelligent techniques are suitable for devising optimal watermarking schemes.

To our knowledge, this chapter is the first survey which classifies and analyses the digital watermarking techniques based on computational intelligence approaches. In order to better understand this complex trend, through the chapter we focus on the core CI approaches, like artificial neural networks and fuzzy sets. Each research work was briefly descried and compared in order to identify the challenges and the potential trend of research. Furthermore, we were able to reveal the advantages and limitations of the existing approaches.

We are aware of data hiding techniques, which make use of evolutionary commutation or chaos theory, or soft computing. In particular, evolutionary computation and soft computing which combine these approaches in order to enhance their advantages, could better secure the multimedia contents. Therefore, a detailed performance analysis of these schemes is another research direction that deserves to be investigated.

References

1. Moulin P, O'Sullivan JA (2003) Information-theoretic analysis of information hiding. IEEE Trans Inf Theory 49(3):563–593
2. Cox IJ, Miller ML (2002) The first 50 years of electronic watermarking. EURASIP J Appl Signal Process 1:126–132
3. Cox IJ, Miller ML, Bloom JA (2001) Digital watermarking. Morgan Kaufmann, pp 26–85
4. Langelaar G, Setyawan I, Lagendijk RL (2000) Watermarking digital image and video data: a state-of-the-art overview. IEEE Signal Process Mag 17:20–46
5. Premaratne P, Ko CC (1999) A novel watermark embedding and detection scheme for images in DFT domain. In: Proceedings of 7th international conference on image processing, London, pp 780–783
6. Cox IJ, Kilian J, Leighton FT, Shamoon T (1997) Secure spread spectrum watermarking for multimedia. IEEE Trans Image Process 6(1):673–1687
7. Inoue H, Miyazaki A, Katsura T (2000) Wavelet-based watermarking for tamper proofing of still images. In: Proceedings of international conference on image processing, pp 88–91
8. Barni M, Bartolini F, Cappellini V, De Rosa A, Piva A (2002) Color image watermarking in the Karhunen-Loève transform domain. J Electron Imaging 11:87–95
9. Ho ATS, Zhu X, Guan YL, Marziliano P (2004) Slant transform watermarking for textured images. In: Proceedings of the 2004 international symposium on circuits and systems, vol. 5, Vancouver, Canada, 23–26 May 2004. IEEE Press, pp 700–703
10. Dugelay JL, Roche S (1999) Fractal transform based large digital watermark embedding and robust full blind extraction. In: Proceedings of IEEE international conference on multimedia computing and systems, pp 1003–1004
11. Basso A, Bergadano F, Cavagnino D, Pomponiu V, Vernone A (2009) A novel block-based watermarking scheme using the SVD transform. Algorithms 1(2):46–75
12. Tsai C-F, Yang W-Y (2007) Real-time color image watermarking based on D-SVD scheme. In: Mery D, Rueda L (eds) Advance in image and video technology, vol 4872. Lecture notes in computer science. Springer, Berlin, pp 289–297
13. Hernandez JR, Amado M, Perez-Gonzalez F (2000) DCT domain watermarking techniques for still images: detector performance analysis and a new structure. IEEE Trans Image Process 9(1):55–68
14. Barni M, Bartolini F, Rosa AD, Piva A (2001) A new decoder for the optimum recovery of non-additive watermarks. IEEE Trans Image Process 10(5):755–766
15. Barni M, Bartolini F, Furon T (2003) A general framework for robust watermarking security. Signal Process J 83:2069–2084
16. Briassouli A, Strintzis MG (2004) Locally optimum nonlinearities for DCT watermark detection. IEEE Trans Image Process 13(12):1604–1607
17. Briassouli A, Sakalides PT, Stouraitis A (2005) Hidden messages in heavy tails: DCT domain watermark detection using alpha-stable models. IEEE Trans Multimedia 7(4):700–715
18. Nikolaidis A, Pitas I (2002) Optimal detector structure for DCT and subband domain watermarking. In: Proceedings of the IEEE international conference on image processing, pp 465–467
19. Voloshynovskiy S, Herrigel A, Baumgärtner N, Pun T (2006) A stochastic approach to content adaptive digital image watermarking. In: Goos G, Hartmanis J, van Leeuwen J

(eds) Proceedings of international workshop on information hiding, Dresden, Germany, 29 Sept–1 Oct 1999. Lecture notes in computer science, vol 1768. Springer, Berlin, pp 211–236

20. Hassanien A-E, Abraham A, Kacprzyk J, Peters JF (2008) Applying genetic programming to evolve learned rules for network anomaly detection. In: Hassanien A-E, Abraham A, Kacprzyk J (eds) Computational intelligence in multimedia processing: recent advances, vol 96. Lecture notes in computer science. Springer, Berlin, pp 3–49
21. Mitra S, Acharya T (2003) Soft computing. In: Mitra S, Acharya T (eds) Data mining: multimedia, soft computing, and bioinformatics. Wiley, pp 35–60
22. Engelbrecht AP (2002) Computational intelligence: an introduction. Willey
23. Bezdek JC (1994) What is computational intelligence? In: Computational intelligence imitating life. IEEE Press, New York, pp 1–12
24. Yardimic A (2009) Soft computing in medicine. Appl Soft Comput 9(3):1029–1043
25. Fulcher J (2008) Computational intelligence: an introduction. In: Fulcher J, Jain LC (eds) Computational intelligence: a compendium, vol 115. Studies in computational intelligence. Springer, Berlin, pp 3–78
26. Hassanien A-E, Abraham A, Kacprzyk J, Peters JF (2008) Computational intelligence: foundation and trends. In: Hassanien A-E, Abraham A, Kacprzyk J (eds) Computational intelligence in multimedia processing: recent advances, vol 96. Studies in computational intelligence. Springer, Berlin, pp 3–49
27. Wu SX, Banzhaf W (2010) The use of computational intelligence in intrusion detection systems: a review. Appl Soft Comput 10(1):1–35
28. Hassoun MH (1995) Fundamentals of artificial neural networks. MIT Press
29. Hertz J, Krogh A, Palmer RG (1991) Introduction to the theory of neural computing. Addison–Wesley Publishing Company
30. Yu P, Tsai H, Lin J (2001) Digital watermarking based on neural networks for color images. Sig Process 81(3):663–671
31. Kutter M, Jordan F, Bossen F (1998) Digital watermarking of color images using amplitude modulation. J Electron Imaging 7(2):326–332
32. Lu W, Lu H, Shen R (2004) Color image watermarking based on neural networks. In: Yin F, Wang J, Guo C (eds) Advances in neural networks, vol 3147. Lecture notes in computer science. Springer, Berlin, pp 651–656
33. Tang G, Liao X-F (2004) A neural network based blind watermarking scheme for digital images. In: Yin F, Wang J, Guo C (eds) Advances in neural networks, vol 3147. Lecture notes in computer science. Springer, Berlin, pp 645–651
34. Zhang J, Wang N-C (2003) Neural network based watermarking for image authentication. J Comput Aided Des Comput Graph 3:307–312
35. Sang J, Alam MS (2005) A neural network based lossless digital image watermarking in the spatial domain. In: Wang J, Liao X, Yi Z (eds) Advances in neural networks, vol 3497. Lecture notes in computer science. Springer, Berlin, pp 772–776
36. Goljan M, Fridrich JJ, Du R (2001) Distortion-free data embedding for images. In: Moskowitz IS (ed) Information hiding, vol 2137. Lecture notes in computer science. Springer, Berlin, pp 27–41
37. Sang J, Liao X, Alam MS, Neural-network-based zero-watermark scheme for digital images. Opt Eng 45
38. Lian S (2009) Image authentication based on neural networks. http://arxiv.org/abs/0707.4524. Accessed 16 Dec 2009
39. Piao C-R, Fan W-Z, Woo D-M, Han S-S (2006) Robust digital image watermarking using BP neural networks. In: Wang J et al (eds) Advances in natural computation, vol 3973. Lecture notes in computer science. Springer, Berlin, pp 285–292
40. Chen B, Wornell GW (2001) Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Trans Inf Theory 4:1423–1443

41. Mei S, Li R, Dang H, Wang Y (2002) Decision of image watermarking strength based on artificial neural networks. In: Proceedings of the 9th international conference on neural information processing, vol. 5, pp 2430–2434
42. Tsai H-H, Cheng J-S, Yu P-T (2003) Audio watermarking based on HAS and neural networks in DCT domain. EURASIP J Appl Sig Process 2003(3):252–263
43. Swanson MD, Zhu B, Tewfik A, Boney L (1998) Robust audio watermarking using perceptual masking. Sig Process 66(3):337–355
44. Lu W, Lu H, Chung F-L (2005) Subsampling-based robust watermarking using neural network detector. In: Wang J, Liao X, Yi Z (eds) Advances in neural networks, vol 3497. Lecture notes in computer science. Springer, Berlin, pp 801–806
45. Chu WC (2003) DCT-based image watermarking using subsampling. IEEE Trans Multimedia 5:34–38
46. Khan A, Tahir SF, Majid A, Coi T-S (2008) Machine learning based adaptive watermark decoding in view of anticipated attack. Pattern Recogn 41(8):2594–2610
47. Birney KA, Fischer TR (1995) On the modeling of DCT and subband image data for compression. IEEE Trans Image Process 4(2):186–193
48. Xu X-Q, Wen X-B, Li Y-Q, Quan J-J (2007) A new watermarking approach based on neural network in wavelet domain. In: Huang D-S, Heutte L, Loog M (eds) Advanced intelligent computing theories and applications: with aspects of artificial intelligence, vol 4682. Lecture notes in computer science. Springer, Berlin, pp 1–6
49. Li D, Wang D, Chen FX (2006) Robust watermark algorithm based on the wavelet moment modulation and neural network detection. J Comput Appl 26:1833–1835
50. Wang D, Li D, Yan J (2008) Robust watermark algorithm based on the wavelet moment modulation and neural network detection. In: Sun F et al (eds) Advances in neural networks, vol 5264. Lecture notes in computer science. Springer, Berlin, pp 392–401
51. Fan Y-C, Mao W-L, Tsao H-W (2003) An artificial neural network-based scheme for fragile watermarking. In: Proceedings of the international conference for consumer electronics, vol 8, pp 210–211
52. Zhao J, Zhou M-Q (2004) A novel wavelet image watermarking scheme combined with chaos sequence and neural network. In: Yin F, Wang J, Guo C (eds) Advances in neural networks, vol 3147. Lecture notes in computer science. Springer, Berlin, pp 663–668
53. Zhao J, Zhao Q, Zhou M-Q, Pan J (2005) A novel wavelet watermark algorithm based on neural network image scramble. In: Wang L, Chen K, Ong YS (eds) Advances in natural computation, vol 3611. Lecture notes in computer science. Springer, Berlin, pp 346–351
54. Lu Y, Han J, Kong J, Yang Y, Hou G (2006) A novel color image watermarking method based on genetic algorithm and hybrid neural networks. In: Greco S et al (eds) Rough sets and current trends in computing, vol 4259. Lecture notes in computer science. Springer, Berlin, pp 806–814
55. Elbasi E, Eskicioglu AM (2006) Neural network based transformation selection in video watermarking. In: Proceedings of the Western New York image processing workshop, Rochester Institute of Technology, 29 Sept 2006, Rochester, NY, USA
56. Zhang ZM, Li RY, Wang L (2003) Adaptive watermark scheme with RBF neural networks. In: Proceedings of the IEEE international conference on neural networks and signal processing, vol. 2, pp 1517–1520
57. Lu W, Lu H, Chung F-L (2006) Robust image watermarking using RBF neural network. In: Wang J et al (eds) Advances in neural networks, vol 3972. Lecture notes in computer science. Springer, Berlin, pp 623–628
58. Piao C-R, Beack S, Woo D-M, Han S-S (2006) A blind watermarking algorithm based on HVS and RBF neural network for digital image. In: Jiao L et al (eds) Advances in natural computation, vol 4221. Lecture notes in computer science. Springer, Berlin, pp 493–496
59. Zhang Y (2009) Blind watermarking algorithm based on HVS and RBF neural network in DWT domain. WSEAS Trans Comput 8(1):174–183

60. Li X, Wang R, A video watermarking scheme based on 3D-DWT and neural network. In: Proceedings of the Ninth IEEE international symposium on multimedia, Taichung, Taiwan, R.O.C.
61. Katzenbeisser S, Petitcolas FAP (eds) Information hiding techniques for steganography and digital watermarking. Artech House, pp 102–103
62. Cheddad A, Condell J, Curran K, Kevitt PM (2010) Digital image steganography: Survey and analysis of current methods. Sig Process 90(3):727–752
63. Orlandi C, Piva A, Barni M (2007) Oblivious neural network computing via homomorphic encryption. EURASIP J Inf Secur 1-11, Article ID 37343
64. Chang Y-C, Lu C-J (2005) Oblivious polynomial evaluation and oblivious neural learning. Theoret Comput Sci 341(1–3):39–54
65. Gorman RP, Sejnowski TJ (1988) Analysis of hidden units in a layered network trained to classify sonar targets. Neural Netw 1(1):75–89
66. Barni M, Orlandi C, Piva A (2006) A privacy-preserving protocol for neural-network-based computation. In: Proceedings of the 8th multimedia and security workshop, Geneva, Switzerland, pp 146–151
67. Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353
68. Kerre EE, Nachtegael M (2000) Fuzzy techniques in image processing: techniques and applications. In: Kerre EE, Nachtegael M (eds) Studies in fuzziness and soft computing, vol 52. Springer, Berlin
69. Tobias OJ, Seara R (2002) Image segmentation by histogram thresholding using fuzzy sets. IEEE Trans Image Process 11(12):1457–1465
70. di Martino F, Loia V, Perfilieva I, Sessa S (2008) An image coding/decoding method based on direct and inverse fuzzy transforms. Int J Approximate Reasoning 48(1):110–131
71. Nachtegael M, Van-Der-Weken D, Van-De-Ville M, Kerre EE, Philips W, Lemahieu I (2001) An overview of classical and fuzzy-classical filters for noise reduction. In: Proceeding of 10th international IEEE conference on fuzzy systems FUZZ-IEEE, Melbourne, Australia, pp 3–6
72. Painter T, Spanias A (2000) Perceptual coding of digital audio. Proc IEEE 88(4):451–513
73. Delay S (1993) The visible difference predictor: an algorithm for the assessment of image fidelity. In: Watson A (ed) Digital image and human vision, chapter 14. MIT Press, pp 212–236
74. Watson AB (ed) (1993) Digital images and human vision. MIT Press
75. Motwani M, Harris Jr FC (2009) Fuzzy perceptual watermarking for ownership verification. In: Proceedings of the 2009 international conference on image processing, computer vision, and pattern recognition, Las Vegas, Nevada
76. Motwani M, Motwani RC, Harris Jr FC (2009) Wavelet based perceptual mask for images. In: Proceedings of IEEE international conference on image processing, Cairo, Egypt
77. Tao B, Dickinson B (1997) Adaptive watermarking in DCT domain. In: Proceedings of IEEE international conference on acoustics, speech and signal processing, vol. 14, no. 4, pp 1985–1988
78. Tao B, Dickinson B, Peterson HA (2000) Adaptive model-driven bit allocation for MPEG video coding. IEEE Trans Circuits Syst Video Technol 10(1):147–157
79. Mohanty SP, Ramakrishnan KR, Kankanhalli MS (2000) An adaptive DCT domain visible watermarking technique for protection of publicly available images. In: Proceedings of the international conference on multimedia processing and systems. Chennai, India, pp 19–25
80. Kankanhalli MS, Ramakrishnan KR (1999) Adaptive visible watermarking of images. In: Proceedings of the IEEE international conference on multimedia computing and systems, vol 1. Florence, Italy, pp 568–573
81. Huang J, Shi YQ (1998) Adaptive Image watermarking scheme based on visual masking. IEE Electron Lett 34(8):748–750
82. Lou DC, Yin T-L, Chang M-C (2004) Robust digital watermarking using fuzzy inference technique. Int J Creative Commun Innov Technol 32(2) (2004)
83. Coumou DJ, Athimoottil M (2001) A fuzzy logic approach for digital image watermarking. In: Proceedings of the Western New York image processing workshop

84. Coumou DJ, Athimoottil M (2004) A fuzzy system solution for digital image watermarking. SPIE Proc 5200:192–196
85. Lee CH, Lee YK (1999) An adaptive digital image watermarking technique for copyright protection. IEEE Trans Consum Electron 45(4):1005–1015
86. Hsieh M-S, Tseng D-C (2005) Multiresolution image watermarking using fuzzy inference filter. In: Proceedings of computational intelligence, Alberta, Canada
87. Hsieh M-S (2001) Wavelet-based image watermarking and compression. PhD dissertation, Institute of Computer Science and Information Engineering, National Central University Chung-li, Taiwan
88. Chen W-C, Wang M-S (2009) A fuzzy c-means clustering-based fragile watermarking scheme for image authentication. Expert Syst Appl 36(2):1300–1307
89. Peng H, Wang J, Wang W (2009) Adaptive image watermarking approach based on kernel clustering and HVS. In: Di Gesu V, Pal SK, Petrosino A (eds) Fuzzy logic and applications, vol 5571. Lecture notes in computer science. Springer, Berlin, pp 213–220
90. Chang C-Y, Wang H-J, Pan S-W (2009) A robust DWT-based copyright verification scheme with fuzzy ART. J Syst Softw 82(11):1906–1915
91. Chang C-Y, Wang H-J, Pan S-W (2007) Robust image hashing scheme based on DWT and fuzzy ART. Technical report, Feng Chia University. http://dspace.lib.fcu.edu.tw/handle/2377/3654Z. Accessed 4 Dec 2009

Chapter 16
Methods to Detect Cyberthreats on Twitter

Praveen Rao, Charles Kamhoua, Laurent Njilla and Kevin Kwiat

Abstract Twitter is a microblogging service where users can post short messages and communicate with millions of users instantaneously. Twitter has been used for marketing, political campaigns, and during catastrophic events. Unfortunately, Twitter has been exploited by spammers and cybercriminals to post spam, spread malware, and launch different kinds of cyberattacks. The ease of following another user on Twitter, the posting of shortened URLs in tweets, the use of trending hashtags in tweets, and so on, have made innocent users the victims of various cyberattacks. This chapter reviews recent methods to detect spam, spammers, cybercus content, and suspicious users on Twitter. It also presents a unified framework for modeling hreats on Twitter are discussed, specifically in the context of big data and adversarial machine learning.

Keywords Cyberthreats · Twitter · Markov logic networks
Modeling and reasoning · Probabilistic inference

Approved for Public Release; Distribution Unlimited: 88ABW-2017-1553, dated 05 Apr 2017.

P. Rao (✉)
Department of Computer Science and Electrical Engineering,
University of Missouri-Kansas City, Kansas City, MO, USA
e-mail: raopr@umkc.edu

C. Kamhoua · L. Njilla · K. Kwiat
Cyber Assurance Branch, Air Force Research Lab, Rome, NY, USA
e-mail: charles.kamhoua.1@us.af.mil

L. Njilla
e-mail: laurent.njilla@us.af.mil

K. Kwiat
e-mail: kevin.kwiat@us.af.mil

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_16

Introduction

In 2016, there were 2.3 billion active users of social media [1]. Among the different social media sites, Twitter, launched in 2006, is a microblogging service where users can post short messages (a.k.a. tweets) and communicate with millions of users instantaneously [2]. According to Twitter, there were 313 million monthly active users in June 2016 [2] generating millions of tweets per day. In recent years, Twitter has become a vital service for politicians to campaign for elections and communicate directly with voters and other outlets. This was clearly witnessed in the 2014 European Parliament elections as well as the 2016 U.S. presidential election. Interestingly, people are able to respond to the politicians' tweets and express their opinions right away.

In addition to politics, social media sites like Twitter are widely used for other activities such as marketing and advertising, sharing breaking news, and during a catastrophic event like an earthquake or a tsunami. Unfortunately, these sites have been exploited by adversaries to launch cyberattacks against users and their organizations [3]. Hackers have taken control of government officials' accounts and also posted false information from popular accounts leading to havoc and economic damages once the posts went viral. In 2012, it was reported in the U.S. media that a rumor was posted on Twitter that the Syrian President had been killed. This caused oil prices to spike [4]. Later in 2015, it was reported again in the U.S. media that the Twitter accounts of U.S. military officials were hacked by those claiming to support the Islamic State [5]. The same year, a U.S. media outlet reported that Iranian hackers took control of social media accounts of U.S. State Department officials [6]. Most recently, it was reported that the Twitter account of Sony Music was hacked and a fake tweet was posted that Britney Spears had died [7].

Another major threat on social media is the spread of malware through social media posts by tricking innocent users to click unsuspecting links. An example is the Koobface malware that spread on Twitter and other social media sites [8]. In 2013, it was reported that a malware infected Twitter users and began collecting users' authentication tokens [9]. Social media is posing an increasing risk for security in the corporate sector [10]. Thus protecting users and systems against cyberthreats on social media has become a critical aspect of enterprise security management.

This chapter provides an overview of recent methods to detect cyberthreats on Twitter. The remainder of this chapter is organized as follows: section "Background" provides a background on the structure and content of publicly available tweets. Section "Cyberthreats on Twitter" discusses cyberthreats on Twitter including techniques for detecting spam and cybercriminals and analyzing the propagation of malware, and techniques for detecting malicious content and suspicious users on Twitter. Section "Future Challenges" presents future challenges in developing cyberthreat detection techniques on Twitter in the context of big data and adversarial machine learning. Finally, a brief conclusion of this chapter is provided in section "Conclusion".

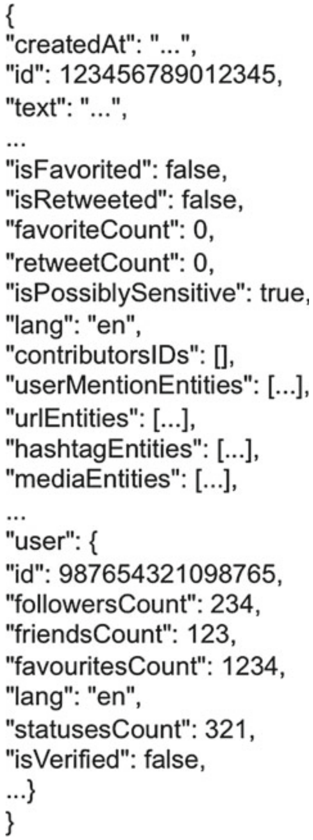
A preliminary version of this work has appeared in the 2nd IEEE International Symposium on Security and Privacy in Social Networks and Big Data (SocialSec) 2016 [11].

Background

A tweet is a 140-character message posted by a user on Twitter. A user can allow his/her tweets to be public. In such a case, anyone can see the tweets posted by the user. Also, anyone can follow a user whose tweets are public without approval from the user. A popular user can thus have a very large number of followers. For example, the account *@realDonaldTrump*, the personal account of the 45th U.S. President, had about 36 million followers in June 2017. During the same month, singer Katy Perry became the first Twitter user to have 100 million followers. When a user posts a tweet, his/her followers will also receive the tweet in their accounts. A user can re-post (a.k.a. retweet) a public tweet to share the information with his/her followers. A tweet that is retweeted more often indicates its growing popularity among users. A user can like a tweet as well as mention other users (as *@username*) in his/her tweets. A user can specify hashtags (prefixed by #) in tweets to indicate specify topics or categories. There are thousands of hashtags in use today. A user can also embed URLs in tweets. Given the 140-character limit of a tweet, shortened URLs typically appear in tweets. Twitter can automatically shorten URLs posted by users, and these begin with https://t.co. Twitter checks the original URLs against suspicious/malicious URLs and and can warn users. But a user can also post shortened URLs using services like Bitly [12] in tweets. This creates a conduit for malicious users to post harmful URLs and trick innocent users to clicking them.

Twitter provides REST APIs for developers to download public tweets in JSON format [13]. Each downloaded tweet is rich in information and very heterogeneous as it may contain 100+ attributes. New attributes may appear over time. A skeleton of a tweet in JSON is shown in Fig. 16.1. (Only certain attributes of the tweet are shown for simplicity.) Each tweet is assigned a unique ID and has a creation time. The attribute *text* contains the actual text typed by the user. The attribute *isFavorited* indicates whether the tweet has been liked by someone. The attribute *isRetweeted* indicates whether the tweet has been retweeted by someone. The attributes *favoriteCount* and *retweetCount* indicate how many users have liked the tweet and retweeted the tweet, respectively. The attribute *isPossiblySensitive* indicates whether Twitter has flagged the URLs (e.g., pointing to media files) in the tweet as sensitive or not. Twitter allows users to report content that is sensitive according its media policy. URLs and hashtags contained in the text of a tweet are extracted and stored under attributes *hashtagEntities* and *urlEntities*, respectively. Similarly, users mentioned in the tweet using *@username* appear under the attribute *userMentionEntities*.

Nested within each tweet is the information about the user who posted the tweet. Each user account is assigned a unique ID. In subsequent discussions, the terms “a user” and “a user account” will be used interchangeably to mean the same thing.



```
{
"createdAt": "...",
"id": 123456789012345,
"text": "...",
...
"isFavorited": false,
"isRetweeted": false,
"favoriteCount": 0,
"retweetCount": 0,
"isPossiblySensitive": true,
"lang": "en",
"contributorsIDs": [],
"userMentionEntities": [...],
"urlEntities": [...],
"hashtagEntities": [...],
"mediaEntities": [...],
...
"user": {
"id": 987654321098765,
"followersCount": 234,
"friendsCount": 123,
"favouritesCount": 1234,
"lang": "en",
"statusesCount": 321,
"isVerified": false,
...}
}
```

Fig. 16.1 Skeleton of a tweet in JSON

For a given user, the attribute *followersCount* indicates how many users are following the user. Twitter defines a friend as someone who a user is following. Hence, the attribute *friendsCount* indicates how many users the given user is following. The attribute *favouritesCount* indicates how many tweets have been liked/favorited by the user. The attribute *statusesCount* indicates the number of tweets posted by the user. Finally, the attribute *isVerified* indicates whether Twitter has verified the user or not. Twitter independently verifies user accounts that are of public interest in domains such as government, fashion, music, politics, sports, and so on. The list of friends or followers of a user is not contained in a downloaded tweet. Nor does the tweet contain information about hashtags that are trending or popular at any point in time. Trending hashtags provide a easy way for users to know what topics are being discussed frequently on Twitter. Note that the list of friends and followers, and trending hashtags can be fetched using Twitter REST APIs.

Cyberthreats on Twitter

This section provides a review of recent methods to detect spam, spammers, cybercriminals, malicious content, and suspicious users on Twitter. It also presents a recent approach called SocialKB [11] to model and reason about the veracity of tweets to discover malicious content and suspicious users on Twitter.

Spam and Spammers on Twitter

While email spam has been around for over two decades, social media sites are increasingly being targetted by spammers. Social spam[1] can be of different types, namely, commercial spam to advertise products, those containing links to phishing and malwares sites, those indicating aggressive behavior such as threats and insults, and so on. Several research attempts have been made to detect spam on Twitter, which are discussed next.

Lee et al. [14] studied the problem of identifying whether a user is legitimate or a spammer based on his/her user profile on Twitter. (They also considered MySpace [15].) To monitor the activity of spammers, social honeypots were set up. Several types of spammers were identified such as those who post duplicate messages to users by mentioning them using @*username* in the tweets, pornographic spammers, those posting phising URLs, and other seemingly legitimate users who post pornographic and commercial content. These social honeypots were able to identify the behavior of spammers, which differed from legitimate users. By training different machine learning based classifiers on spam and legitimate profiles, new spam accounts on Twitter were detected. About one-fifth of the spam accounts were identified as bots that posted tweets automatically.

Stringhini et al. [16] studied the problem of detecting spammers and spam campaigns on Twitter by creating 300 honey-profiles on Twitter. (They also considered Facebook [17] and MySpace.) They collected the messages (tweets and direct messages) and requests received by these profiles for a period of 11 months between 2009 and 2010. After careful analysis of the collected data, they used features based on the number of followers and friends of a profile, the ratio of messages sent by a profile containing URLs, similarity of messages sent by a profile, and so on, to train a classifier based on Random Forest. The accuracy of the results were tested by sending the detected spam profiles to Twitter.

Grier et al. [18] observed that less than one-tenth of the URLs posted on Twitter were spam URLs. They found that spammers used mentions, retweets, trending hashtags, tweet hijacking, etc., to target users. Therefore, they developed techniques to identifiy spam campaigns based on the blacklisted landing pages posted by spam accounts. They showed that URL blacklists were slow in flagging spam URLs, and hence spammers were successful in tricking users to click these URLs before they

[1] https://en.wikipedia.org/wiki/Social_spam.

were blacklisted. Further, the use of shortened URLs made the detection of spam URLs more challenging.

With the growing use of shortened URLs on Twitter, Wang et al. [19] analyzed the misuse of shortened URLs in tweets. They gathered over 600,000 Bitly's shortened URLs and classified them into spam and non-spam based on the click traffic data. Specifically, they considered features based on user and global-level clicks, distribution of clicks from different countries, and clicks from referrers (*i.e.*, web pages or applications that contain the shortened URLs). Among the different classifiers, they showed that Random Tree algorithm achieved the best accuracy.

Like the Web, Twitter is also vulnerable to link farming by users. Ghosh et al. [20] studied this issue, wherein one can attempt to acquire a large number of followers. By increasing the number of followers, one can increase their influence in the network. The study showed that spammers attempted to increase the number of followers by following other users and expecting a fraction of them to follow them back. The study found that unlike the Web, link farming was mainly done by a small number of legitimate, popular, and active users. A ranking scheme was proposed to dissuade link farming by such users (for following spammers) by penalizing them via lower influence scores. As a result, this would deter spammers from increasing their social influence in the network.

Cybercriminals and Spreading of Malware on Twitter

Yang et al. [21] analyzed Twitter accounts to understand how criminal accounts mesh with users on Twitter. Specifically, they analyzed the inner social relationships of criminal accounts, and how these accounts connected with other users (*i.e.*, outer social relationships), who support these criminal accounts. They concluded that criminal accounts tend to be connected as a small-world network. The hubs in the criminal network were more likely to follow more criminal accounts than the leaf nodes in the network. Among the supporters of criminal accounts, two prominent types were identified: those that simply follow others that follow them, and those who wish to promote their business/products by following others without careful consideration. Finally, an inference algorithm for identifying criminal accounts was proposed by leveraging a seed set of known criminal accounts and social relationships and similarity of the tweets posted.

Sanzgiri et al. [22] developed and studied different attack models to analyze the spread of malware on Twitter. The attacks included simple attacks such as when an attacker posts shortened malicious URLs and uses @*username* to target random users, or takes control of a legitimate account and posts malicious URLs to target the followers of the account. More complex attacks included clickjacking and inserting trending hashtags in tweets with malicious URLs. Via simulation results, the authors showed that even with low degree of connectivity with users and low probability of users clicking the posted links, an attacker can infect several users.

Lee et al. [23] developed a system for detecting suspicious URLs on Twitter streams. They show that attackers can host malicious content in websites that are reachable via several redirects from the initially posted (shortened) URLs in tweets. When crawlers attempt to investigate these malicious links, they are conditionally redirected to non-malicious landing pages.[2] This makes it harder to identify these links as malicious ahead of time. However, if the correlated URL redirect chains are considered, then certain entry point URLs used by the attackers can be identified. The system considered different classifiers based on the features from correlated URL redirect chains and tweet context information. It was shown that the classifier based on L2-regularized logistic regression yielded the best accuracy.

Recently, Burnap et al. [24] developed a classification system using machine learning to detect malicious URLs in a matter of seconds after the links are clicked. They collected tweets posted during two popular sporting events. Tweets from one event was used to train a classifier. The tweets from the other event was used for testing the classifier. Using a client-side honeypot, they collected machine activity logs when URLs were clicked. The honeypot was used to flag malicious URLs based on changes to registry, file system, running processes, and so on. Different classifiers were trained based on the behavior of malicious software using the machine activity logs. The features used were based on the CPU usage, bytes sent and received, packets sent and received, and others. Among the different classifiers, a Multi-Layer Perceptron-based approach achieved the best accuracy on the testing data.

A Unified Framework for Modeling and Reasoning About the Veracity of Twitter Data

The techniques discussed in sections "Spam and Spammers on Twitter" and "Cybercriminals and Spreading of Malware on Twitter" were designed to detect specific threats on Twitter such spam, spammers, cybercriminals, malicious content, and suspicious users. However, none of these techniques provide a unified framework to model and reason about the veracity of posts of Twitter and user behavior all within the same framework. Motivated by these reasons, SocialKB was developed by Rao et al. [11] to discover suspicious users and malicious content. SocialKB relies on advances in statistical relational learning to construct a knowledge base (KB) over Twitter data so that both the behavior of users and the nature of their posts can be analyzed within the same framework, by reasoning over the constructed KB via probabilistic inference. As a result, SocialKB provides the flexibility to model and methodically analyze several of the key attributes in tweets, which other techniques may fail to consider.

[2] For example, http://google.com.

Markov Logic Networks

In statistical relational learning, a Markov logic network (MLN) [25] is regarded as one of the most flexible representations as it combines first-order logic and probabilistic graphical models. First-order logic enables the complexity of the data to be modeled; and probability allows expressing the uncertainty in the data. MLNs are widely used in natural language processing [26], entity resolution [27, 28], hypertext classification [29], and information extraction [30] and retrieval. Formally, a MLN is a KB defined by a set of pairs (F,w), where F is a first-order formula that denotes a constraint and w is a real-valued weight of the formula. Higher the weight, more likely is the constraint believed to be satisfied in a possible world. A formula with a positive weight is more likely to be satisfied in a possible world; a formula with a negative weight is more likely to be unsatisfied. A formula with infinite weight is a hard constraint. Formulas in the KB can contradict. A world that violates a formula is less probable but not impossible. However, a world that violates a hard constraint has zero probability. A grounding of a formula (or predicate) is obtained by replacing all its variables by constants. The obtained formula (or predicate) is called a ground formula (or ground predicate).

Once the formulas and weights are learned [31], probabilistic inference can be performed on the MLN by posing maximum a posteriori (MAP) and marginal inference queries. A MAP query outputs the most likely world (e.g., a set of ground predicates that are most likely to be satisfied). A marginal inference query outputs the marginal probabilities (e.g., of ground predicates). Efficient inferencing techniques have been developed in recent years such as lifted inference [32, 33] as well as those that leverage the scalability and efficiency of relational database systems and cluster computing (*e.g.*, Tuffy [34], ProbKB [35]). As a result, it is now possible to reason on large KBs with millions of entities, facts, and relationships.

Example 1 Consider a simple KB about Twitter users. Let *attacker*(u) denote a predicate that is true if user u is an attacker and false otherwise. Let *verified*(u) denote a predicate that is true if u is a verified user and false otherwise. Consider the 3 formulas: $\forall u$ *attacker*(u); $\forall u$ *verified*(u); and $\forall u$ *verified*(u) $\implies$!*attacker*(u) with weights −2.0, 1.0, and 3.0, respectively. The first formula with negative weight implies that a user is more likely to be a non-attacker. Compared to the first two formulas, the third formula is a stronger constraint and is of higher importance in the set of possible worlds. Using probabilistic inference, one can perform marginal inference queries on a ground predicate or all ground predicates (*e.g.*, $\mathrm{Pr(attacker(Alice))}$, $\mathrm{Pr(attacker(x))}$) as well as MAP queries (*e.g.*, $\arg\max_x \mathrm{Pr(attacker(x))}$). □

Given a MLN, its ground Markov network is denoted by (X, G), where X is the set of binary random variables and G is the set of ground formulas. For each ground predicate, there is one binary random variable in X. The set of possible worlds $\mathscr{X}$ is the set of all possible assignments of truth values to variables in X. The probability of a possible world is given by $\mathrm{Pr(X = x) = \frac{1}{Z} \exp(\sum_i w_i n_i(x))}$, where w_i is the weight of the ith formula, $n_i(x)$ is the number of true groundings of the ith formula in x, and $Z = \sum_{x' \in \mathscr{X}} \exp(\sum_i w_i n_i(x'))$ is a normalization constant.

Predicates in SocialKB

SocialKB has different types of predicates in the KB, which evaluate to `true` or `false` after grounding. A predicate can make a closed-world assumption (CWA) or an open-world assumption (OWA). CWA assumes that what is not known to be true must be false. On the other hand, OWA assumes that what is not known may or may not be true.

Table 16.1 shows the first set of predicates in the KB based on CWA. The variables in these predicates, namely, *tweetID*, *userID*, *link*, and *hashtag* denote the ID of a tweet, the ID of a user, a URL, and a hashtag, respectively. The predicate *tweeted(userID,tweetID)* is used to indicate that a particular tweet was posted by a particular user; *containsLink(tweetID,link)* is used to indicate that a tweet contains a particular URL; *mentions(tweetID,userID)* is used to indicate that a particular user is mentioned in a tweet (using the @ symbol); *retweeted(userID,tweetID)* is used to indicate that a user retweeted a particular tweet; *containsHashtag(tweetID,hashtag)* is used to indicate that a tweet contains a particular hashtag; finally, *verified(userID)* is used to indicate that a user has been verified by Twitter.

Table 16.1 also shows a set of predicates based on OWA. The predicate *malicious(link)* is used to indicate that a URL is malicious; *friend(userID1,userID2)* is used to indicate that a user denoted by *userID1* has a friend denoted by *userID2*; *trending(hashtag)* is used to indicate that a hashtag is trending; *attacker(userID)* is used to indicate that a user is a suspicious user/attacker; *isFollowedBy(userID1, userID2)* is used to indicate that a user denoted by *userID1* is followed by another user denoted by *userID2*; and finally, *isPossiblySensitive(tweetID)* is used to indicate that a tweet is possibly sensitive.

To model the count information in a tweet as well as about users' social relationships, a set of predicates are defined as shown in Table 16.2. These predicates are used to indicate the friends count, followers count, and statuses count of a user, retweet count of a tweet, and the number of tweets a user has liked/favorited. These

Table 16.1 Predicates to capture relationships based on tweet content and user information

Predicate	Type	Predicate	Type
tweeted(userID,tweetID)	CWA	malicious(link)	OWA
containsLink(tweetID,link)	CWA	friend(userID1,userID2)	OWA
mentions(tweetID,userID)	CWA	trending(hashtag)	OWA
retweeted(userID,tweetID)	CWA	attacker(userID)	OWA
containsHashtag(tweetID,hashtag)	CWA	isFollowedBy(userID1,userID2)	OWA
verified(userID)	CWA	isPossiblySensitive(tweetID)	OWA

Table 16.2 Predicates to model counts in a tweet and temporal events

Predicate	Type	Predicate	Type
friendsCount(userID,count)	CWA	tweetedT(userID,tweetID,t)	CWA
followersCount(userID,count)	CWA	trendingT(hashtag,t)	OWA
statusesCount(userID,count)	CWA	followersCountT(userID,count,t)	CWA
retweetCount(tweetID,count)	CWA	friendsCountT(userID,count,t)	CWA
favouritesCount(userID,count)	CWA	favouritesCountT(userID,count,t)	CWA

predicates make a CWA. Table 16.2 also shows predicates with temporal variables, which enables SocialKB to model sequence of events over time. The variable *t* denotes the temporal attribute. The predicate *tweetedT(userID, tweetID, t)* is used to indicate when a user posted a particular tweet; *trendingT(hashtag, t)* is used to indicate when a hashtag is trending; *followersCountT(userID, count, t)* indicates when a user has a particular followers count; and finally, *friendsCount(userID, count, t)* indicates when a user has a particular friends count. Similar temporal predicates can be defined on retweet count and statuses count.

Formulas in SocialKB

In the current version of SocialKB, the first-order formulas of the KB were constructed based on findings in published literature [19, 21, 22, 24], observing account activities on Twitter, and through intuitive reasoning. Figure 16.2 shows all the formulas. Some of the formulas in the KB can be thought of as untested hypotheses. Their true effect on the results of probabilistic inference will depend on their actual weights, which will be learned from the evidence dataset and queries. The existential quantifier ∃ on each variable in a formula is implied.

The first set of formulas ($\mathbf{f_1} - \mathbf{f_2}$) infers a friend relation between two users based on mentions and retweets. The second set of formulas ($\mathbf{f_1} - \mathbf{f_2}$) infers a trending hashtag based on its usage by an attacker and a verified/legitimate user when he/she is followed by a verified user. The third set of formulas ($\mathbf{f_5} - \mathbf{f_{10}}$) infers whether a user is an attacker/suspicious user or not. Essentially, a verified user implies that he/she is not an attacker; a friend of a verified user who also follows the user is not an attacker; a user who posts a malicious URL is an attacker; a user who is a friend of an attacker and is followed by the attacker is an attacker; a user who is mentioned by another legitimate/trustworthy user is also trustworthy; a user whose tweet is possibly sensitive is an attacker.

The fourth set of formulas ($\mathbf{f_{11}} - \mathbf{f_{13}}$) infers whether a URL is malicious or not. These formulas imply that a URL with a trusted domain name is not malicious (e.g., https://t.co), a URL contained in a possibly sensitive tweet is malicious, and a URL in a tweet posted by an attacker is malicious. The fifth set of formulas ($\mathbf{f_{14}} - \mathbf{f_{15}}$) infers a possibly sensitive tweet if a tweet containing a malicious URL or is from an attacker.

$\mathbf{f_1}$: tweeted(userID1,tweetID) $\wedge$ mentions(tweetID,userID2) $\Longrightarrow$ friend(userID1,userID2)

$\mathbf{f_2}$: retweeted(userID1,tweetID) $\wedge$ tweeted(userID2,tweetID) $\Longrightarrow$ friend(userID1,userID2)

$\mathbf{f_3}$: tweeted(userID,tweetID) $\wedge$ containsHashtag(tweetID,hashtag) $\wedge$ attacker(userID) $\Longrightarrow$ trending(hashtag)

$\mathbf{f_4}$: isFollowedBy(userID1,userID2) $\wedge$ verified(userID2) $\Longrightarrow$ verified(userID1)

$\mathbf{f_5}$: verified(userID) $\Longrightarrow$!attacker(userID)

$\mathbf{f_6}$: verified(userID1) $\wedge$ friend(userID1,userID2) $\wedge$ isFollowedBy(userID1,userID2) $\Longrightarrow$!attacker(userID2)

$\mathbf{f_7}$: tweeted(userID,tweetID) $\wedge$ containsLink(tweetID,link) $\wedge$ malicious(link) $\Longrightarrow$ attacker(userID)

$\mathbf{f_8}$: attacker(userID1) $\wedge$ friend(userID1,userID2) $\wedge$ isFollowedBy(userID1,userID2) $\Longrightarrow$ attacker(userID2)

$\mathbf{f_9}$: !attacker(userID1) $\wedge$ tweeted(userID1,tweetID) $\wedge$ mentions(tweetID,userID2) $\Longrightarrow$!attacker(userID2)

$\mathbf{f_{10}}$: tweeted(userID,tweetID) $\wedge$ isPossiblySensitive(tweetID) $\Longrightarrow$ attacker(userID)

$\mathbf{f_{11}}$: containsLink(tweetID,link) $\wedge$ [contains(link,ϕ)] $\Longrightarrow$!malicious(link) // hard constraint with wt. ∞

$\mathbf{f_{12}}$: containsLink(tweetID,link) $\wedge$ isPossiblySensitive(tweetID) $\Longrightarrow$ malicious(link)

$\mathbf{f_{13}}$: attacker(userID) $\wedge$ tweeted(userID,tweetID) $\wedge$ containsLink(tweetID,link) $\Longrightarrow$ malicious(link)

$\mathbf{f_{14}}$: containsLink(tweetID,link) $\wedge$ malicious(link) $\Longrightarrow$ isPossiblySensitive(tweetID)

$\mathbf{f_{15}}$: attacker(userID) $\wedge$ tweeted(userID,tweetID) $\Longrightarrow$ isPossiblySensitive(tweetID)

$\mathbf{f_{16}}$: !verified(user) $\wedge$ followersCount(user,count1) $\wedge$ friendsCount(user, count2) $\wedge$ [count1 != 0 AND count2/count1 $> m_1$] $\Longrightarrow$ attacker(user)

$\mathbf{f_{17}}$: !verified(user) $\wedge$ statusesCount(user,count1) $\wedge$ friendsCount(user,count2) $\wedge$ [count1 != 0 AND count2/count1 $> m_2$] $\Longrightarrow$ attacker(user)

$\mathbf{f_{18}}$: !verified(user) $\wedge$ statusesCount(user,count1) $\wedge$ followersCount(user,count2) $\wedge$ [count1 != 0 AND count2/count1 $> m_3$] $\Longrightarrow$ attacker(user)

$\mathbf{f_{19}}$: !verified(user) $\wedge$ statusesCount(user,count1) $\wedge$ favouritesCount(user,count2) $\wedge$ [count1 != 0 AND count2/count1 $> m_4$] $\Longrightarrow$ attacker(user)

$\mathbf{f_{20}}$: friendsCountT(user,count1,t1) $\wedge$ friendsCountT(user,count2,t2) $\wedge$ [t2 - t1 <= τ AND count1 != 0 AND count2/count1 $> m_5$] $\Longrightarrow$ attacker(user)

$\mathbf{f_{21}}$: trendingT(hashtag,t1) $\wedge$ tweetedT(user,tweet,t2) $\wedge$ containsHashtag(tweet,hashtag) $\wedge$ containsLink(tweet,link) $\wedge$ attacker(user) $\wedge$ [t1 < t2] $\Longrightarrow$ malicious(link)

$\mathbf{f_{22}}$: trendingT(hashtag,t1) $\wedge$ tweetedT(user1,tweet,t2) $\wedge$ mentions(tweet,user2) $\wedge$!isFollowedBy(user2,user1) $\wedge$ [t1 < t2] $\Longrightarrow$ attacker(user1)

$\mathbf{f_{23}}$: trendingT(hashtag,t1) $\wedge$ tweetedT(user1,tweet,t2) $\wedge$ mentions(tweet,user2) $\wedge$!friend(user2,user1) $\wedge$ [t1 < t2] $\Longrightarrow$ attacker(user1)

Fig. 16.2 First set of first-order formulas in the KB

The sixth set of formulas ($\mathbf{f_{16}} - \mathbf{f_{19}}$) infers attackers/suspicious users based on the counts of certain attributes in the tweets. In these formulas, $m_1, \ldots, m_4$ denote positive integer constants. These formulas imply that a non-verified user is an attacker if the user is following a very large number of users compared to the number of users who are following the user, and that if a non-verified user is not active on Twitter (based on the number of tweets posted) but has a large number of friends/followers or has liked a large number of tweets, then the user is an attacker. These trends are indications that a user wants to increase his/her social influence. Note that when a user's tweet is liked by someone, then a notification is sent to the user. Thus, an attacker can drawn the attention of other users to himself/herself by randomly liking their tweets. Similarly, a user can mention another user in his/her tweet to seek attention to posts that could contain potentially malicious content.

The last set of formulas ($\mathbf{f_{20}} - \mathbf{f_{23}}$) is defined over predicates with temporal variables. These formulas are powerful to model a sequence of activities, which can be exploited by adversaries to launch cyberattacks. For example, if the friends count of a user increases substantially during a predefined time interval τ, then the user is an attacker as he/she is trying to increase his/her social influence rapidly. If a hashtag is trending at a point in time, and an attacker posts a tweet containing that hashtag at a later time, and if the tweet contains a URL, then it is implied to be malicious. This captures the actions of an attacker who is tracking trending hashtags to post malicious URLs to maximize the scope of an attack. Also if a hashtag is trending at a point in time, and a user posts a tweet containing that hashtag at a later time, and mentions another user who he/she is not following or is not friends with, then the user is an attacker. This constraint models attackers who can mention other users in their posts randomly just to have malicious content sent to those innocent users.

Architecture of SocialKB

The overall architecture of SocialKB is shown in Fig. 16.3. It uses Tuffy [34] as the underlying MLN framework. A key component of SocialKB is the KB containing first-order predicates and formulas with both hard and soft constraints. In the current version of SocialKB, the formulas were handcrafted. (However, it is possible to learn new formulas automatically by applying rule mining techniques.) Given the KB, the first step is to generate an evidence dataset. SocialKB uses external data sources such as URLBlacklist.com to mark malicious URLs. This evidence dataset contains ground predicates in the KB of the MLN that are known to be satisfied. The next step is to learn the weights of the formulas in the KB given a set of queries for which inference will be performed. For example, *malicious(link)*, *attacker(userID)*, and *isPossiblySensitive(tweetID)* are examples of inference queries. The final step is to perform probabilistic inference on the set of queries using the learned MLN. The output of the MAP and marginal inference queries can be verified using Twitter REST APIs and services such as VirusTotal [36], a free online service that aggregates the output/analysis of several antivirus engines and website scanners on suspicious URLs and files.

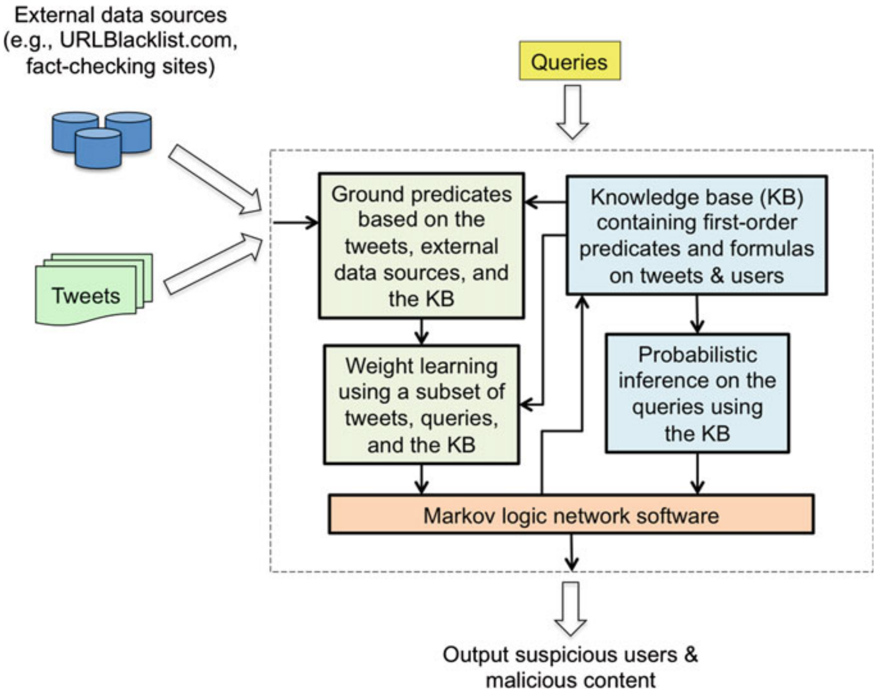

External data sources
(e.g., URLBlacklist.com,
fact-checking sites)
Tweets
Queries
Ground predicates based on the tweets, external data sources, and the KB
Knowledge base (KB) containing first-order predicates and formulas on tweets & users
Weight learning using a subset of tweets, queries, and the KB
Probabilistic inference on the queries using the KB
Markov logic network software
Output suspicious users & malicious content

Fig. 16.3 Overall architecture

Evaluation of SocialKB

Rao et al. reported a preliminary evaluation of SocialKB on 20,000 tweets [11]. They considered three queries: *attacker(userID)*, *malicious(link)*, and *isPossiblySensitive(tweetID)*. They used URLBlacklist.com to mark URLs as malicious in the evidence dataset. VirusTotal was used to validate the URLs after the inference task. A URL was flagged as malicious if it was reported as malicious by at least one antivirus engine/scanner on VirusTotal using direct URL search or after resolving the IP address of the URL's domain and fetching the aggregate report. For *malicious(link)*, about 84% of the URLs output by MAP inference were associated with malicious content. For *attacker(userID)*, about 80% of the users output by the MAP inference had posted URLs that were malicious either based on URLBlacklist.com or using VirusTotal. Some of the user accounts were suspended by Twitter. Finally, for *isPossiblySensitive(tweetID)*, 95% of the tweets output by the MAP inference contained URLs that were flagged as malicious by URLBlacklist.com or using VirusTotal. Similar trends were seen for the output of marginal inference on these three queries when the output ground predicates with probability 0.8 or higher were analyzed.

While the initial results obtained by SocialKB appear promising, it remains to be seen how well SocialKB can perform compared to classifiers based on logistic regression and Random Forests. Another important area of investigation is the automatic learning of first-order formulas of the KB on large-scale Twitter data. An efficient approach is necessary to cope with the evolving behavior of users and nature of content posted on Twitter.

Future Challenges

While machine learning has been frequently used to develop techniques for detecting cyberthreats on Twitter, several future challenges arise in the context of big data and adversarial machine learning, which are discussed in this section.

Coping with Big Data

Social media datasets can be regarded as *big data* due to their massive, complex, heterogeneous nature. According to a White House report [37], big data analytics will be transformative in the area of national security and can provide resilience to cyberattacks. However, the volume, velocity, and veracity of big data pose new challenges for detecting cyberthreats on Twitter and other social media sites. Many of the techniques proposed for detecting threats on Twitter apply machine learning techniques. One challenge in the context of big data is obtaining large labeled datasets to accurately to build machine learning models for supervised learning. This could be time-consuming and require a lot of manual effort. Thus, semi-supervised learning techniques, which make use of a small number of label data and a much larger number of unlabeled data during learning, can be explored for massive social media datasets. Another challenge is the scalability of machine learning techniques to tackle massive structured and semi-structured datasets. Fortunately, new frameworks are being developed to scale machine learning algorithms using cluster computing [38–40]. While this solves part of the problem, it is still a challenge as to how often the models should be retrained as new data arrive. This would be necessary when users' behavior and nature of posts change over time leading to a change in the underlying distribution of the data.

Adversarial Machine Learning

Adversarial machine learning is the study of machine learning techniques in the presence of adversaries, who have the capability and resources to mislead these techniques [41]. It is gaining a lot of attention in security as machine learning techniques are popularly used in intrusion detection systems, spam filters, and virus scanners. The techniques discussed in this chapter for detecting cyberthreats on Twitter must be studied in the presence of sophisticated attackers. Unlike intrusion detection systems, virus scanners, and spam filters, which have matured over the years, social media sites are still in their infancy w.r.t. thwarting cyberattacks. A few possible attack scenarios are discussed next. One possible attack is when an attacker can influence the training data so as to mislead the learner. For example, malicious users could pose as legitimate users for certain duration and posts tweets and create social relationships that are genuine/harmless. These posts could be used by the learner during training/retraining leading to incorrect decisions. Another way is that the attacker can continuously try different actions and learn about the outcome of the action by obtaining statistics of the clicks to shortened URLs, whether an account got suspended, or tweets got flagged and so on. The cyberthreat detection techniques based on machine learning must make suitable assumptions about the knowledge of an attacker. The attacker may have full knowledge of the learning algorithm, and in some cases, the actual features used during training. However, the attacker may not have complete information about the data used for training and evaluation.

Several vulnerabilities exist during learning, which can be exploited by adversaries [41]. Assumptions such as data linearity, feature independence, data stationarity where the training and testing data are assumed to come from the same distribution, and the presence of independent and identically distributed (i.i.d.) data points make the learning algorithms susceptible to attacks. One strategy to tackle malicious training data is to use the Reject on Negative Impact (RONI) defense technique [42]. If adding a new training instance significantly affects the classification accuracy, then that instance is rejected as being malicious. Another strategy is to leverage techniques from robust statistics, which can tolerate a small amount of malicious data/outliers during learning [41]. It remains to be seen how cyberthreat detection techniques for Twitter can be developed that are robust to attacks by adversaries.

Conclusion

While Twitter connects millions of people instantaneously via short messages, it is also an indispensable information source for many outlets. On Twitter, users can follow other users with a simple click, post shortened URLs (in tweets) using different services, mention other users in tweets, use any hashtags, and so on. Unfortunately, spammers and cybercriminals have found different ways to exploit the simplicity and flexibility of features on Twitter causing different kinds of cyberattacks. This chapter discussed various methods and techniques to detect spam, spammers, cybercriminals, malicious content, and suspicious users on Twitter. Many of the techniques discussed in this chapter apply machine learning techniques by considering user behavior, content of tweets, social relationships, etc., to detect different types of cyberthreats. It also discussed a unified framework called SocialKB for modeling tweets and detecting suspicious users and malicious content on Twitter via probabilistic inference. Unlike prior techniques, SocialKB can analyze both the behavior of users and the nature of their posts on Twitter–all within the same framework. In the context of big data and adversarial machine learning, this chapter also discussed several challenges with regards to applying machine learning for detecting cyberthreats on Twitter.

While this chapter did not discuss the growing problem of fake news/rumors on social media, it must be noted that detection of such cyberthreats deserves immediate attention.

Acknowledgements This work was performed while the first author held an NRC Research Associateship award at Air Force Research Lab, Rome, New York. The authors would like to thank the anonymous reviewers for their comments and suggestions, and Anas Katib for his assistance.

References

1. Digital in 2016. http://wearesocial.com/special-reports/digital-in-2016, Jan 2016
2. Twitter Usage. https://about.twitter.com/company, Feb 2017
3. Cisco (2013). https://www.cisco.com/web/offer/gist_ty2_asset/Cisco_2013_ASR.pdf
4. Twitter Death Rumor Leads to Spike in Oil Prices. http://mashable.com/2012/08/07/twitter-rumor-oil-price, Aug 2012
5. FBI Investigating Central Command Twitter Hack. http://www.cnbc.com/2015/01/12/us-central-command-twitter-hacked.html, Jan 2015
6. Iranian Hackers Attack State Dept. via Social Media Accounts. https://www.nytimes.com/2015/11/25/world/middleeast/iran-hackers-cyberespionage-state-department-social-media.html, Nov 2015
7. Sony Music's Twitter Hacked, Fake Britney Spears Death Tweets Sent. http://www.reuters.com/article/us-sony-twitter-cyber-idUSKBN14F11D, Dec 2016
8. Thomas K, Nicol DM (2010) The Koobface botnet and the rise of social malware. In: Proceedings of the 5th International conference on malicious and unwanted software, Oct 2010, pp 63–70
9. Twitter Malware: Spreading More Than Just Ideas. https://securityintelligence.com/twitter-malware-spreading-more-than-just-ideas, Apr 2013
10. Social Media a Growing Risk for Corporate Security (2016). https://gdssummits.com/app/uploads/sites/1/2016/03/Social-media-a-growing-risk-for-corporate-security-whitepaper.pdf
11. Rao P, Katib A, Kamhoua C, Kwiat K, Njilla L (2016) Probabilistic inference on Twitter data to discover suspicious users and malicious content. In: Proceedings of the 2nd IEEE International symposium on security and privacy in social networks and big data (SocialSec), Nadi, Fiji, pp 1–8
12. Bitly (2017). https://bitly.com
13. Twitter Developer Documentation (2017). https://dev.twitter.com/rest/public
14. Lee K, Caverlee J, Webb S (2010) Uncovering social spammers: social honeypots + machine learning. In: Proceedings of the 33rd International SIGIR conference, pp 435–442
15. Myspace (2017). https://myspace.com
16. Stringhini G, Kruegel C, Vigna G (2010) Detecting spammers on social networks. In: Proceedings of the 26th annual computer security applications conference, pp 1–9
17. Facebook (2017). https://www.facebook.com
18. Grier C, Thomas K, Paxson V, Zhang M (2010) @Spam: the underground on 140 characters or less. In: Proceedings of the 17th ACM conference on computer and communications security, Chicago, Illinois, USA, pp 27–37

19. Wang D, Navathe SB, Liu L, Irani D, Tamersoy A, Pu C (2013) Click traffic analysis of short URL spam on Twitter. In: Proceedings of 9th International conference on collaborative computing: networking, applications and worksharing, Oct 2013, pp 250–259
20. Ghosh S, Viswanath B, Kooti F, Sharma NK, Korlam G, Benevenuto F, Ganguly N, Gummadi KP (2012) Understanding and combating link farming in the Twitter social network. In: Proceedings of the 21st International conference on world wide web, pp 61–70
21. Yang C, Harkreader R, Zhang J, Shin S, Gu G (2012) Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on Twitter. In: Proceedings of the 21st International conference on the world wide web, pp 71–80
22. Sanzgiri A, Hughes A, Upadhyaya S (2013) Analysis of malware propagation in Twitter. In: Proceedings of the 32nd IEEE symposium on reliable distributed systems, pp 195–204
23. Lee S, Kim J (2013) WarningBird: a near real-time detection system for suspicious URLs in Twitter stream. IEEE Trans Dependable Secur Comput 10(3):183–195
24. Burnap P, Javed A, Rana OF, Awan MS (2015) Real-time classification of malicious URLs on Twitter using machine activity data. In: Proceedings of the 2015 IEEE/ACM International conference on advances in social networks analysis and mining 2015, pp 970–977
25. Richardson M, Domingos P (2006) Markov logic networks. Mach Learn 62(1–2):107–136
26. Poon H, Domingos P (2008) Joint unsupervised coreference resolution with Markov logic. In: Proceedings of the conference on empirical methods in NLP, pp 650–659
27. Mccallum A, Wellner B (2004) Conditional models of identity uncertainty with application to noun coreference. In: Saul LK, Weiss Y, Bottou L (eds) Advances in neural information processing systems 17. MIT Press, Cambridge, MA, pp 905–912
28. Singla P, Domingos P (2006) Entity resolution with Markov logic. In: Proceedings of the 6th International conference on data mining, ICDM '06, pp 572–582
29. Chakrabarti S, Dom B, Indyk P (1998) Enhanced hypertext categorization using hyperlinks. In: Proceedings of the 1998 ACM SIGMOD International conference on management of data, Seattle, Washington, USA, pp 307–318
30. Poon H, Domingos P (2007) Joint inference in information extraction. In: Proceedings of the 22nd national conference on artificial intelligence—volume 1, Vancouver, British Columbia, Canada, pp 913–918
31. Singla P, Domingos P (2005) Discriminative training of Markov logic networks. In: Proceedings of the 20th AAAI conference on artificial intelligence, pp 868–873
32. Jha AK, Gogate V, Meliou A, Suciu D (2010) Lifted inference seen from the other side: the tractable features. In: Proceedings of advances in neural information processing systems (NIPS), pp 973–981
33. Sarkhel S, Singla P, Gogate V (2015) Fast lifted MAP inference via partitioning. In: Proceedings of advances in neural information processing systems (NIPS), pp 3240–3248
34. Niu F, Ré C, Doan A, Shavlik J (2011) Tuffy: scaling up statistical inference in Markov logic networks using an RDBMS. Proc VLDB Endow 4(6):373–384
35. Chen Y, Wang DZ (2014) Knowledge expansion over probabilistic knowledge bases. In: Proceedings of the 2014 ACM SIGMOD conference, pp 649–660
36. VirusTotal (2017). https://virustotal.com
37. Big Data: Seizing Opportunities, Preserving Values (2014). http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf
38. Li M, Andersen DG, Park JW, Smola AJ, Ahmed A, Josifovski V, Long J, Shekita EJ, Su B-Y (2014) Scaling distributed machine learning with the parameter server. In: 11th OSDI conference, Oct 2014, pp 583–598
39. Low Y, Gonzalez J, Kyrola A, Bickson D, Guestrin C, Hellerstein JM (2012) Distributed GraphLab: a framework for machine learning in the cloud. Proc. VLDB Endow 5(8):716–727
40. Meng X, Bradley JK, Yavuz B, Sparks ER, Venkataraman S, Liu D, Freeman J, Tsai DB, Amde M, Owen S, Xin D, Xin R, Franklin MJ, Zadeh R, Zaharia M, Talwalkar A (2015) MLlib: machine learning in Apache Spark. CoRR. arXiv:1505.06807
41. Huang L, Joseph AD, Nelson B, Rubinstein BI, Tygar JD (2011) Adversarial machine learning. In: Proceedings of the 4th ACM workshop on security and artificial intelligence, Chicago, Illinois, USA, pp 43–58

42. Nelson B, Barreno M, Jack Chi F, Joseph AD, Rubinstein BIP, Saini U, Sutton C, Tygar JD, Xia K (2009) Misleading learners: co-opting your spam filter. Springer US, Boston, MA, pp 17–51

Chapter 17
Detecting Users Who Share Extremist Content on Twitter

Yifang Wei and Lisa Singh

Abstract Identifying extremist-associated conversations on social media sites and blog forums is still an open problem. Extremist groups leverage social media to (1) spread their message and (2) gain recruits. In this chapter, we look at different work in this arena, focusing on metrics and features that researchers have proposed as proxies for *misbehavior* on Twitter. We begin this chapter by analyzing potential features a small amount of manually labeled data about ISIS supporters on Twitter. We then group these features into categories related to tweet content, viewpoints, and dynamics. After discussing different state of the art methods for extremism detection and similar problems, we present a case study looking at the ISIS extremist group. Finally, we discuss how one collects these data for a surveillance system and conclude by discussing some current challenges and future directions for effective surveillance of extremism.

Introduction

As Twitter, Facebook, and other social media sites continues to grow in popularity, conversations involving extremism are being recognized as a serious problem. One well known extremist group using Twitter as a platform for sharing its ideas and recruiting members/jihadists to its group is the Islamic State of Iraq and Syria (ISIS), also known as IS or ISIL. ISIS is itself responsible for mass atrocities in Iraq and Syria, and the group uses Twitter to encourage supporters to initiate terrorist attacks worldwide [1]. Its activities have also led to large-scale displacement in the Middle East. At of the end of 2015, almost 6 million persons were internally displaced in

Y. Wei (✉) · L. Singh
Georgetown University, Washington, DC, USA
e-mail: yw255@georgetown.edu

L. Singh
e-mail: lisa.singh@georgetown.edu

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_17

Syria and another 4 million were refugees in neighboring countries. ISIS related attacks in Iraq had displaced approximately 2.6 million persons in that country [2, 3].

To reduce extremist conversation by individuals and terrorist groups, Twitter has been suspending accounts which are believed to be associated with terrorist organizations. In February 2016, Twitter announced that it had shut down 125,000 accounts related to ISIS between the middle of 2015 to the beginning of 2016 [4]. The removed ISIS-related accounts include ISIS-related media outlets, information hubs, and supporters. For obvious reasons, Twitter has never released its algorithm or strategy for determining whether an account is primarily related to ISIS or not [5]. In this chapter, we consider a generalization of this problem-determining whether a social media account should be classified as exhibiting extremist behavior, e.g. sharing extremist content, or not. We will refer to this problem as *extremism detection*. More formally, given a set of social media publishers U, our goal is to identify the subset of publishers U^+ that exhibit extremist behavior by promoting or disseminating extremist content.

The chapter is organized as follows. Section "Extremism Detection Case Study" presents a simple analysis of tweets containing extremist content and accounts exhibiting extremist behavior. Given this content, section "Relevant Features for Extremism Detection" considers different features that may be useful for identifying these accounts exhibiting extremist behavior on Twitter. In section "State of the Art" we overview recent results and state of the art systems. Section "Twitter Data Analysis" presents a case study focusing on detecting extremism related to ISIS on Twitter. Finally, section "Surveillance Systems for Extremism Detection" presents suggestions for data collection to gather these relevant data set for surveillance and concludes by presenting directions and challenges for our research community.

Twitter Data Analysis

In this section we present a small case study describing extremist related data on Twitter. We discuss both tweet content and extremist accounts.

We begin by looking at some examples of extremist content. Table 17.1 shows examples of some tweets that were classified as supporting an extreme viewpoint. We see that these tweets are retweeted by a small fraction of individuals who exhibit extremist behavior compared to the overall number of individuals who retweet them. When designing algorithms that attempt to identify extremism, we need to consider both content and content propagation. What are the distinguishing words? What are the frequent ones? Words like *honor*, *win*, and *love* are not unusual and will be used in many positive contexts as well. To further complicate the situation, extremist content will be propagated by both extremists and non-extremists, so just following the flow of messages is insufficient.

To better understand content on Twitter, we manually identify approximately 1300 individual tweets that exhibit extremist views consistent with views of the

Table 17.1 Examples of extremist tweets and their retweet propagation

Tweet	Extremist retweets	Overall retweets
These 2 pics will explain to you the meaning of the whole world,1: Jihad = Honor 2: Democracy = Humiliation	3	162
This is how #ISIS supporters in Twitter win against western world :) :) http://t.co/...	3	154
See the difference between orphanages in #ISIS and other countries..thats why we love #IslamicState...	2	34

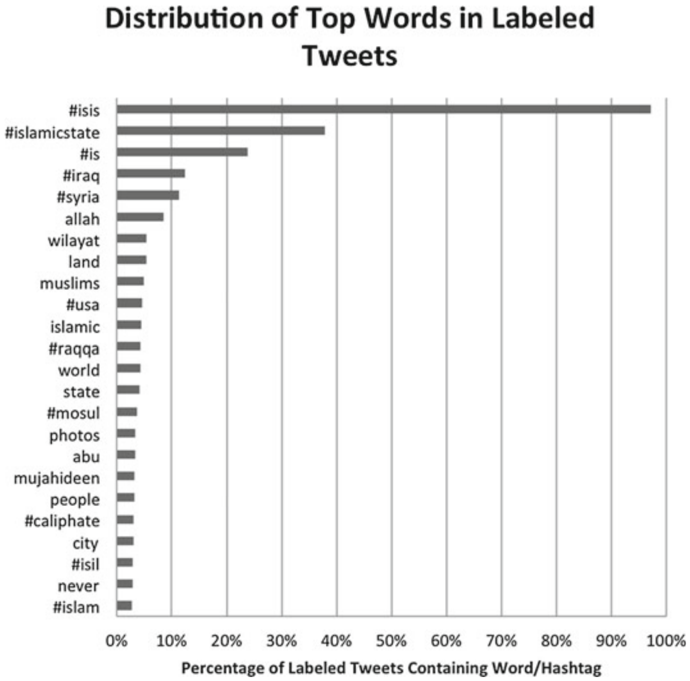


Fig. 17.1 Distribution of Words Across Tweets

ISIS extremist group. Figure 17.1 shows the frequency distribution of the different words as a percentage of the total number of tweets. As expected, the predominant words in these tweets are the hashtags of the extremist group itself, i.e. synonyms for ISIS [`#isis` (97%), `#islamicstate` (38%), `#is` (24%) and `#isil` (3%)]. The next group of popular words are related to religion, specifically Islam [`Muslim` (66%), `Islamic` (59%), and `#islam` (37%)]. There are general words like `world`, `state`, `people`, `never`, that are in 35–60% of the tweets. Another set of words are related to locations in Iraq and Syria, as well as general geography words, e.g. `land` also appear regularly in these tweets. `Iraq` (12%) and `Syria` (11%) are the most frequent. Interestingly, the `United States` occurs in 5% of these tweets.

Finally, there are a small group of extremist words that occur in approximately 3% of the tweets [Abu (the name of an extremist), Mujahideen (a person engaged in Jihad), and #caliphate (an Islamic state led by a religious leader who is considered a successor to the Islamic prophet Muhammad.)] It is interesting to note that there are very few "extreme" words in this set of tweets. This reminds us that searching for more unique words that are associated with extremist thought would miss a large number of tweets containing extremist content. The bag of words model will help identify some tweets that contain extremist views, but will also miss a large number of tweets because similar vocabulary is used in different contexts. For this reason, other features like sentiment/tone, and stance need to be considered. These features begin to get at the opinion of the post's author.

Finally, we look at five Twitter accounts that exhibit extremist behavior. Some consistent properties exist across these accounts. First, all of these users tweet a fair amount. On average, over 4500 tweets per year. Second, while they have followers, the median number of followers is 942 and the high is only 2166. The network structure when considering follower and friend count are very similar to regular Twitter users. In other words, the overall network structure is not unique when compared to accounts not exhibiting extremist behavior for this set of data. This is one reason both network structure and content of the followers and friends is important-counts alone do not always tell the entire story.

Relevant Features for Extremism Detection

Given this small glimpse into the content on Twitter, we see that there are a large number of features that are relevant for identifying extremist behavior. Unfortunately, social media data is noisy, partial, and biased. This means that there are additional challenges when building classifiers in this arena. In this section, we identify features that can be used to improve the reliability of different algorithms for extremist detection. Later in the chapter we discuss data quality challenges. Note that we focus on Twitter features, but they can easily be generalized to features on other social media sites. We discuss a few interesting, specific examples throughout. We organize the relevant features into the following categories: user post basic content, user post viewpoint, user social media profile, user network profile, user content dynamics, and user dynamics.

User Post Basic Content

User post basic content features focus on identifying distinguishing words, symbols, and sentence structure in posts. The simplest feature is word frequency. Are the words that appear more frequently in posts containing extremist content different from words that appear in posts containing non-extremist content. The vocabulary is also important for determining the topic of a post. People do not express the same ideas the same ways. So it is important to group words into topics to better understand the similarities and differences between the content of different posts.

While one can mine the data to determine the relevant set of words for different extremist groups and topics, custom dictionaries developed by experts in the field are another important source for identifying words, synonyms, and topics that may be relevant. Other tweet specific content features, include URLs, punctuation, capitalization, hashtags, mentions, emoticons, emojis, capitalization, geotagged location, photos, photo captions, and photo metadata. While some of these features may seem useless, e.g. punctuation, each of these features has been useful for machine learning tasks related to understanding behaviors on Twitter (in different contexts). Therefore, it is important to determine which are insightful when learning behavioral norms for certain extremist conversations. Finally, the structure of the tweet can also give insight into its purpose-specific features include, the language usage, the sentence structure, and the tweet readability.

User Post Viewpoint

User post viewpoint features attempt to interpret the tone/sentiment and emotion of the post. Is the post negative or positive toward the particular extremist group or toward extremist actions? There are different opinion or viewpoint features that can be measured including, sentiment/tone, stance, and emotion. Sentiment/tone analysis focuses on whether the post is positive, negative, or neutral. It does not consider opinion in the context of the topic of the tweet. To capture this, a feature that can be measured is *stance* [6]. Stance detection focuses on determining if the post writer is in favor, against, or neutral toward a target. Finally, while sentiment and stance are important, so is emotion. What are the emotions expressed in the post? Anger? Happiness? Despair? Effective content analysis requires that we not only understand the topic of the post, but also its tone, its stance and the emotion expressed by the author of the post. When the user shares a reasonable fraction of posts that have extremist words, negative sentiment, a stance that supports extremist groups or viewpoints, and content that is angry or meant to incite, the user is exhibiting extremist behavior.

User Social Media Profile

The *user social media profile* looks at features specific to a user account. The user's name, affiliation, profile information, social media account name (sometimes these are inflammatory), interest lists, groups joined, and location information, to name a few. These features give us insight into basic demographics and interests of the user. While the features will tend to be less distinguishing, there are some who have very overt extremist behavior that can be identified using these more general features.

User Network Profile

The *user network profile* looks specifically at the composition of the user's local neighborhood on the social media site. In the context of Twitter, features would include the number of followers, the number of followers that retweet extremist content, the number of people the user follows that share extremist content, the size of the follower network compared to others, the clustering coefficient of the user's network (how many of the users followers follow each other), and the sentiment, stance, and emotional profile of the user's network. If the user does not post a large amount of content himself, having this network information can be useful for identifying his potential extremist views.

Just as interesting is considering how the network associations change through time. Are more and more individuals with extremist behavior following a particular user? Is the number decreasing? Is the user beginning to follow other users that are sharing extremist viewpoints. Understanding these changing dynamics of the network are also potentially valuable features for this task.

User Content Dynamics

Extremist content on social media gets varying amounts of attention. *User content dynamics* focuses on features that measure the popularity of content. Is a particular post being discussed and/or retweeted more than the normal post? Are people who are reposting or retweeting it sympathetic to the extremist content in the original post or not? In other words, what is their stance on the post? Here we need to be creative with our feature generation. Possible features include the number of retweets of user content by followers exhibiting extremist features and the stance of those retweeting the content. If the content shows support for extremist ideology or thought, we may see more people denouncing it than supporting it. We also need to understand the accounts being mentioned-perhaps some of those are accounts of individuals sharing extremist content. And finally, the change in topic between the original tweet and the retweets. Are there new themes that are being generated by the tweet, e.g. new hashtags, that people are beginning to add to their own tweets?

User Dynamics

User dynamics focuses on the users distribution frequency of different types of content. Here we focus on developing a profile of the user based on the types of content shared overall (not just a specific post). For a given user, what is the distribution of his post content? What is the distribution of emotion in the content? What types of content does the user retweet? Understanding the distribution of different content streams is important for understanding the similarities and differences between those that share extremist content and those that do not.

State of the Art

In a social media setting, a surveillance system may have access to a restricted set of relevant features. It is unlikely that all the mentioned features will be available. Therefore, it is important to understand what can be determined with different subsets of these features. In this section, we review the state of the art. We highlight the features used and the results obtained in different relevant studies. At the end of the chapter we will discuss the future of surveillance in this field given the data challenges.

Extremism detection is a relatively new area of research in the context of social media data. This is partially because data has not been available to the academic community and earlier extremist groups did not use social media as extensively as it has been used in the last few years. One form of abusive behavior that has been studied more extensively on Twitter is spam detection [7–10]. While relevant, spam propagates differently from extremist content—the campaign style is different. First, a major goal of spam is to increase visibility of a product or idea. It is not recruitment. This means that *conversations* are not needed to have a successful campaign. Second, a spam campaign is usually focused on generating profit. An extremist campaign promotes an ideology and focuses on convincing others to believe in a particular extreme viewpoint. This leads to extremist posts that have more sophisticated content and tone than traditional spam. Finally, spam in social media is usually circulated by robots, while extremist conversations/campaigns are executed by human supporters, voluntarily in most cases. Previous literature has shown that patterns of human behavior in social media differs significantly from that of robots [11]. Given all these reasons, we believe that the behavioral patterns associated with extremist detection will vary significantly from spam detection, and therefore, different algorithms are necessary for accurate extremism detection.

We pause to mention that there are relevant areas of the literature that we do not explore in this chapter. We refer you to surveys in this area that are broader than this piece [12, 13].

Identifying Extremist Content

A primary task of extremist content detection on social media is crawling extremist content, for which several solutions have been proposed [14, 15]. Mei and Frank [14] classify a webpage into four sentiment-based classes: pro-extremist, anti-extremist, neutral, and irrelevant using a sentiment and word frequency based decision tree classifer. They propose a web crawler capable of crawling webpages with pro-extremist sentiment, achieving 80% accuracy. Bouchard et al. [15] identify a set of words that are used to distinguish terrorist websites from anti-terrorist websites using content analysis of different types of websites, e.g. white supremacist websites, jihadist websites, terrorist related news websites, and counter-terrorism websites.

This type of dictionary is important for understanding the goal/mission of these different sites. They used this feature analysis to develop a web crawler which automatically searches the Internet for extremist content. Looking at these examples of the literature in this area, we see that while there are reasonable methods for identifying website extremist content, the features used are fairly basic. As more robust features related to web hyperlinks, stance, and emotion are considered, classification accuracy is likely to continue to improve.

Understanding Spread

Beyond simply crawling extremist contents, [16–18] analyze content exposing extremist ideology. Chatfield et al. [16] investigate the problem of how extremists leverage social media to spread their propaganda. They perform network and content analysis of tweets published by a user previously identified as an information disseminator of ISIS. Burnap et al. [17] study the propagation pattern of the information following a terrorist attack. Zhou et al. [18] analyze the hyperlink structure and the content of the extremist websites to better understand connections between extremist groups. Buntain et al. [19] study the response of social media to three terrorist attacks: the 2013 Marathon bombing in Boston, the 2014 hostage crisis in Sydney, and the 2015 Charlie Hebdo shooting in Paris. Not surprisingly, they found that the use of retweets, hashtags, and urls related to the events increased during and after the events. These different findings reinforce the importance of information dissemination and spread for extremism behavior detection.

Learning from Networks Using Extremist Seeds

Understanding the networks of different extremists is an important direction of research. One approach to doing this is to begin with a set of individuals who have been identified (most likely manually) as affiliated with extremist groups or are aliases of known extremist accounts. As an example, Berger and Morgan [5] start with approximately 400 manually selected Twitter accounts which they believe to be official accounts of ISIS, and use different content and network features of a 2-hop network for each of these accounts to build a classifier that identifies supporters/sympathizers of ISIS. Their classifier had an 80% accuracy on the labeled data. Berger and Strathearn [20] leverage an information flow network to identify accounts serving as information hubs for promoting extremist ideology. They find that if the begin with a few seed individuals, using metrics for influence, exposure and interactivity are sufficient for identifying individuals engaging in extremist conversation.

Learning from Content Without Extremist Seeds

Suppose we are not given a set of seed individuals that are known to post extremist content. This means we do not have knowledge of a user's network in advance and must rely on content shared and the flow of the shared content. Wei and Singh [21] identify extremist behavior by dividing the problem into two subproblems: identifying relevant posts and then using those posts to identify individuals sharing content consistent with extremist views. Because different extremist groups use different vocabulary on social media, generic dictionaries are less effective. Therefore, the authors begin by identifying features that best distinguish seed posts exhibiting extremist ideology from seed posts exhibiting anti-extremist ideology. They then use these distinguishing features to identify relevant posts. The posts are then used to construct two weighted networks that model the information flow between the publishers of the identified posts. Different node centrality metrics are considered for evaluating a users' contribution to spreading extremist ideology and anti-extremist ideology. Their approach leads to an accuracy of 90% for the top extremist users, but the accuracy deteriorates for users who have more limited content and message flow. Rowe and Saif [22] investigate ways to identify behavioral changes in an individual when they transition from a "normal" state to one that is pro-ISIS. Their approach considers changes in retweeting behavior of extremist material and language usage changes, e.g. increased use of keywords that are considered pro-ISIS and/or anti-Western. They find that in a data set of over 150,000 Twitter users, less than 1000 exhibit pro-ISIS behavior.

While progress is being made in this arena, these methods are still in their infancy. They only consider a small fraction of the features we have described and have not effectively calibrated the impact of the different types of features. Future work needs to understand the limits and biases associated with content analysis.

Other Types of Inappropriate Behavior Detection

There are many other threads of relevant research related to detecting inappropriate behavior. These include promoting marijuana [23], attacks of identity theft [24], identifying suspicious urls [25], and finding worms that are propagating [26]. These papers leverage different content, including follower demographics, url reuse, etc. While the profiles built have some of the same features at a broad level, the dictionaries and specific features used must be custom designed. Still, it is important to remember that these other tangential areas give us insight into the subsets of features that are distinguishing in similar domains.

Sentiment and Stance

We pause to discuss sentiment analysis since it is important for understanding tone. Tweet sentiment analysis has been an active area of research in recent years. Most sentiment algorithms are supervised. The commonly employed features include word-related features (words, stems, n-grams) [27–29], word shape features (punctuations, and capitalization) [28], syntactic features (POS taggers, dependency trees) [27, 29, 30], and Twitter-specific features (Twitter handles, hashtags, urls, emojis, emoticons) [27, 28, 30]. Li et al. [31] propose an unsupervised algorithm that uses a sentiment topic model with decomposed priors. Lexicon-based methods do not require training data. Gutierrez et al. [32] propose to classify Spanish sentence sentiment using support vector machines and a Spanish sentiment lexicon. Instead of simply using the fixed sentiment polarity of words, Saif et al. [33] update the sentiment of words based on co-occurrences of words in different contexts. As previously mentioned, Mei and Frank [14] classify a webpage into four sentiment-based classes: pro-extremist, anti-extremist, neutral, and irrelevant. Their sentiment classifier considers only nouns.

Most of the approaches mentioned above are target-independent sentiment analysis. There is no doubt about an occurrence of a complimentary word strongly indicates positive sentiment. However, our task considers features that are needed for target-dependent sentiment analysis, i.e., we aim to classify the sentiment in a tweet towards an extremist group. This is where the idea of stance comes in. Liang et al. [28] define 7 types of syntactic patterns, to extract co-occurrences of a target object and expressions with sentiment. Similarly, Nasukawa et al. [34] adopt a set of human-created rules for nouns, verbs, and adjectives to extract sentiments towards specific objects. Hu et al. [35] summarize sentiment of customer reviews towards products. Since the customer reviews could be considered towards specific products presumably, all the expressed sentiment is relevant to the target products for their task. Finally, Mohammad et al. [6] use a linear-kernel SVM classifier that combines features drawn from training data with more general features to determine stance. Their classification performance is around 80% and they show the difference in sentiment and stance and the need for both.

Extremism Detection Case Study

This section is a summary of work in Wei et al. [3]. We conduct a small exploratory case study using approximately 2 million accounts, identifying the number of accounts exhibiting extremist behavior.

Features for Identifying Extremist Behavior

For this analysis, we consider the following features: sentiment/stance of tweets, the polarity of the user's ego-network, and user mentions as different proxies for misbehavior.

Sentiment/Stance Tendency Feature Anecdotal observations support the idea that a user with extremist views consistent with an extremist group will show positive sentiment towards that group in his/her posts, while an ordinary person will show negative/neutral sentiment towards the group. Based on this observation, our sentiment tendency feature $\mathscr{S}$ measures the sentiment tendency for an individual based on the overall sentiment of his/her published tweets. To create this feature, we begin by identifying the tweets associated with a particular extremist group. One simple approach is to look for any variant of the name of the group in the tweet. For ease of exposition, let T equal the set of tweets that are relevant for a particular user. Each tweet is assigned a sentiment value (positive or negative/neutral). Then $\mathscr{S} = \frac{\sum_{i=1}^{|T|} sentiment(t_i)}{|T|}$, where $sentiment(t_i)$ is the sentiment of tweet t_i. A score of 1 is assigned if the sentiment of a tweet is positive. A score of 0 is assigned if the sentiment of the tweet is neutral or negative. Therefore, if $\mathscr{S} > 0.5$, the sentiment is classified as positive.

Ego-Network Extremism Support Feature We make the following two observations about extremist content on Twitter: (1) a user with extremist views consistent with an extremist group is highly likely to be followed by at least one user exhibiting similar extremist views, and (2) an ordinary user might follow a user exhibiting extremist-associated behaviors. Adversaries, socialists, journalists, researchers, including researchers on our team, do this to monitor/learn behaviors of individuals with extremist views. Based on these two observations, if a user has one or more followers that have a positive sentiment tendency $\mathscr{S} > 0.5$ and the user has a positive sentiment tendency, our ego-network extremism support feature $\mathscr{E} = 1$. Otherwise, it is zero.

Mention-Network Feature When analyzing the data, we observed the following: (1) Users with extremist content in tweets are highly likely to be mentioned by other users with a similar viewpoint, and (2) an ordinary user might mention a user exhibiting extremism-associated behaviors in his/her tweets. Based on these two observations, our mention network feature $\mathscr{M} = 1$ if at least one other user exhibiting extremist sentiment tendencies mentions the user and the user has a positive sentiment tendency.

Data Set Exploration

We collected tweets with references to any of these hashtags between September 2014 and April 2016 using Twitter API. Table 17.2 shows the number of tweets

Table 17.2 The number of tweets containing different ISIS related hashtags

Language	Hashtag	# Tweets
English	#ISIS	17 million
	#ISIL	2.5 million
	#IS	3.2 million
	#Islamic_State	20 thousand
	#Islamic_State_in_Iraq_and_Al-Sham	1.5 thousand
	#Islamic_State_in_Iraq_and_the_Levant	5 thousand
Arabic	ISIS	14 million
	ISIL	44 million
	IS	6.7 million
	Islamic State	25 thousand
	Islamic State in Iraq and the Levant	1.2 million
	Islamic State of Iraq and Al-Sham	46 thousand

associated with the main ISIS related hashtags in both English and Arabic during our data collection period. Tweets containing an English hashtag pertinent to ISIS were published by approximately 2 million Twitter users, and tweets containing an Arabic hashtag pertinent to ISIS were published by approximately 1.2 million Twitter users.

We use the Twitter API to obtain tweets with references to ISIS-relevant hashtags as the first round of data collection. The second round of data collection focuses on collecting profiles of the users who published one or more of the downloaded ISIS-related tweets. Since we began data collection, 11% of the 2 million Twitter users that published tweets using one or more of the English ISIS-related hashtags have been suspended, while 23% of the accounts using the Arabic ISIS-related hashtags have been suspended. These suspensions occurred during our data collection. For those subset of users, we do not have their network information or any tweets after the account suspension.

Tweet Sentiment/Stance Classification

When using a basic sentiment analyzer that considers just content for measuring sentiment and stance, we found that the majority of tweets did not contain any sentiment or stance, i.e. they are more objective or neutral. Based on previous literature, we build a custom stance classifier that contained the following features: unigrams (except stopwords), emoticons, URLs, hashtags, and negation expressions. Our project team containing Middle East experts manually labelled 3800 English tweets using the following categories: positive (positive stance towards ISIS), negative (negative stance towards ISIS), neutral, and noise (tweets considered noise are

Table 17.3 Accuracy of classifiers based on different models and state-of-art sentiment analysis tools in classifying tweet sentiment

Classifier	Accuracy%
NB	85
LR	85
SVM	70
KNN	51
Stanford CoreNLP	68
vaderSentiment	53
SentiStrength	57

removed from the later experiments). Our experts were very specific with their labeling. For example, negative stance towards ISIS is different from negative sentiment in a tweet discussing President Obama's policies related to ISIS.

Since our goal is to detect posts supporting extremist views, a tweet with positive stance towards ISIS is labeled *positive*, while a tweet with negative/neutral stance towards ISIS is labeled *negative*. We consider four different binary classifiers: Naive Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). Table 17.3 shows the accuracies achieved by these classifiers using 10-fold cross validation. We also test three popular sentiment analysis tools (Stanford CoreNLP Sentiment Analyzer [36], vaderSentiment [37], and SentiStrength [38]) on the same dataset. All of these three tools are set in binary mode (positive VS negative/neutral), and their accuracies are shown in Table 17.3. We see that our classifiers (NB, LR, and SVM) outperform these state-of-art tools. Given these results, we use Naive Bayes for the remaining experiments related to this case study.

Extremism Detection

In order to evaluate our extremism detection results, researchers on our project team manually identified 100 accounts with content consistent with pro-ISIS views among suspended accounts, 100 with content consistent with pro-ISIS views among unsuspended accounts, 100 with content consistent with anti/neutral-ISIS views among suspended accounts, and 100 with content consistent with anti/neutral-ISIS among unsuspended accounts. The extremist detection decisions given by different combinations of features presented in the previous subsection were then evaluated against the hand labeled results: sentiment/stance-tendency (SENT), the ego-network extremism support (EGO), and the mentioned-network (MENTIONED). The results are shown in Fig. 17.2. SENT only considers the user's sentiment/stance tendency. EGO only considers the user's ego-network extremism support.

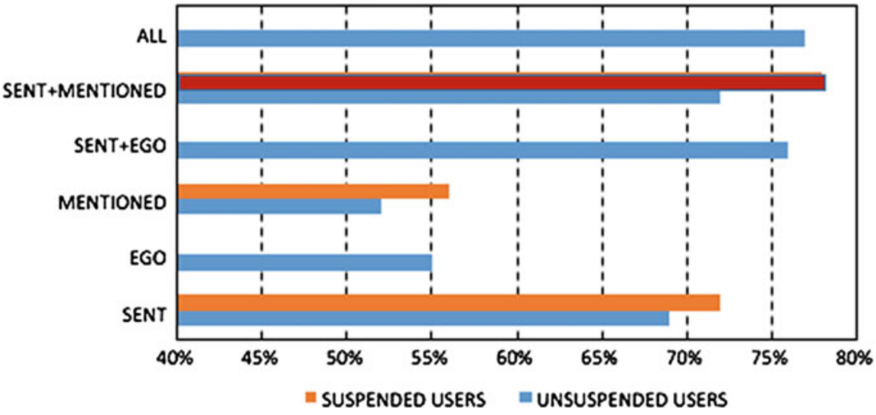


Fig. 17.2 Extremism detecting accuracy by different detectors

Table 17.4 The number of accounts determined as having content consistent with pro-ISIS views

		Suspended accounts	Unsuspended accounts
SENT	Positive	24,761	49,951
	Negative	174,100	1,769,559
EGO	Positive	NA	36,784
	Negative	NA	1,782,726
MENTIONED	Positive	15,832	23,549
	Negative	183,029	1,795,961
SENT+EGO	Positive	NA	42,158
	Negative	NA	1,777,352
SENT+MENTIONED	Positive	18,542	40,352
	Negative	180,319	1,779,158
ALL	Positive	NA	30,671
	Negative	NA	1,788,839

MENTIONED only considers the user's mentioned-network. SENT+EGO considers the user's sentiment/stance tendency and the user's ego-network extremism support. SENT+MENTIONED considers the user's sentiment/stance tendency and the user's mentioned-network. ALL combines all the features. Note that we separate our analysis of the suspended accounts from the not suspended accounts because we do not have EGO features for the suspended accounts. From Fig. 17.2, we see that when classifying the unsuspended users, the detector combining all the three features outperforms the others; for classifying the suspended accounts the detector combining the SENT and the MENTIONED features has the highest accuracy.

Our final analysis was using all the proposed features to detect the number of extremist accounts in our full data set. Table 17.4 shows the number of users classified as having content consistent with pro-ISIS views and anti-ISIS views. While we do not have a ground truth to validate the results, the numbers confirm our intuition that suspended accounts have a much higher percentage of users having content consistent with pro-ISIS views than unsuspended accounts; While misclassification does exist, the proportion of misclassification should be relatively similar in the 2 data sets. Using this approach for surveillance may help Twitter identify these accounts more rapidly.

Surveillance Systems for Extremism Detection

Capturing Data

Surveillance of social media extremist behavior is very challenging. Social media sites limit the amount and type of data they are willing to let people download using their free APIs. This adds a new level of complexity to surveillance. Even in the best case scenario, a group does not have funds to pay for all the data a social media site has. So a feature selection process needs to take place. Here we suggest different options for collecting relevant Twitter data: (1) Identify a small number of Twitter accounts exhibiting extremist behavior manually and download the tweets, followers and friends for those accounts. Those accounts can be used for a small case study, but they can also be considered seed accounts. It is best if the seed accounts are not friends and followers of each other. Once all the available data is collected from these seed accounts, we can begin collecting tweets, profile information and summary network statistics for the users that are friends and followers of these seed individuals. Good collection will continue building these ego networks from this snowball sample until the data collection period is over; (2) Identify hashtags that correspond to the names or ideals of different extremist groups. Use the API to collect data from these streams; (3) Randomly select the names of individuals who post negative content on one of the hashtag related data streams. Conduct a 2-hop snowball sample beginning with these accounts as seeds. Each of the accounts captured needs to be continually collecting data every few weeks. Otherwise, any emerging dynamics or changing behaviors will not be detected.

Keep in mind while this process will result in raw data that is meaningful, many of the features described need to be constructed from these raw data using state of the art methods described in the previous section and by developing new methods as well. Another consideration is the processing power and storage space needed for these types of surveillance systems. We try to keep older data to help us understand the changing dynamics of the extremist groups. It is important to continually update models based on current data. At the same time, the historic data gives us an opportunity to establish ground truth for different predictive tasks. As long as data sets are not available for this purpose for these types of analyses, we must create our own data sets using different social media APIs.

Current Challenges and Future Directions

Extremism detection is a challenging problem, not just algorithmicly, but also politically. Identifying individuals incorrectly can lead to loss of livelihood and reputation for the individual. Not detecting individuals (missing them) can lead to unexpected attacks and further recruitment into extremist groups. These additional considerations remind us of the importance of quality predictions and the need to qualify all of our results with confidence levels. How reliable is the result and are there biases that need to be considered? The more precise we are about the quality and confidence of the results, the more likely it is that we can reduce the potential for misidentification. Finally, surveillance needs to be conducted in a way that preserves the privacy of individuals who share their data. To help mitigate challenges with private data, we focus on conducting surveillance using only public data. Everyone, regardless of viewpoint, deserves to have their private data protected. Without user consent, using private, sensitive data, should not be considered an option.

As a community, if we are going to make progress on these larger societal-scale problems, we need to work together—share algorithms, data sets, and testing platforms. We need to look at the problem from more than a computational perspective. Algorithms need to incorporate changing behaviors by working with social scientists who have studied extremist behavior from psychological, social, and political perspectives. We will not succeed in building realistic models from these noisy, partial data sets without getting expert knowledge from those who have been studying extremism for decades. Extremism has been a problem throughout time; however, social media platforms give those with extremist views a new recruiting tool that cannot be ignored. If we can reduce the impact of this new recruiting tool, we can reduce the overall impact of extremism.

Acknowledgements We would like to acknowledge the Georgetown Institute for the Study of International Migration (ISIM) for their subject matter expertise throughout the process. This work was supported in part by the National Science Foundation (NSF) Grant SMA-1338507 and the Georgetown University Mass Data Institute (MDI). Any opinions, findings, conclusions, and recommendations expressed in this work are those of the authors and do not necessarily reflect the views of NSF or MDI.

References

1. Lister T, Sanchez R, Bixler M, O'Key S, Hogenmiller M, Tawfeeq M (2017) Isis goes global
2. Un high commissioner for refugees (2016) global trends: forced displacement in 2015. http://www.unhcr.org/en-us/statistics/unhcrstats/576408cd7/unhcr-global-trends-2015.html?query=Global%20trends%202016
3. Wei Y, Singh L, Martin S (2016) Identification of extremism on twitter. In: Workshop on social network analysis surveillance technologies (SNAT) at the IEEE/ACM international conference on advances in social network analysis and mining
4. Twitter's new ISIS policy. http://www.theatlantic.com/international/archive/2016/02/twitter-isis/460269/

5. Berger J, Morgan J (2015) The ISIS twitter census: defining and describing the population of ISIS supporters on twitter. Brook Proj US Relat Islam World 3(20)
6. Mohammad SM, Sobhani P, Kiritchenko S (2016) Stance and sentiment in tweets. arXiv:1605.01655
7. Wang AH (2010) Don't follow me: spam detection in twitter. In IEEE SECRYPT. pp 1–10
8. Song J, Lee S, Kim J (2011) Spam filtering in twitter using sender-receiver relationship. In: Recent advances in intrusion detection. Springer, pp 301–317
9. Benevenuto F, Rodrigues T, Almeida V, Almeida J, Gonçalves M (2009) Detecting spammers and content promoters in online video social networks. In: SIGIR. ACM, pp 620–627
10. Lee K, Caverlee J, Webb S (2010) Uncovering social spammers: social honeypots+machine learning. In: SIGIR. ACM, pp 435–442
11. Lee J, Cha S, Lee D, Lee H (2009) Classification of web robots: an empirical study based on over one billion requests. comput secur 28(8):795–802
12. Agarwal S, Sureka A (2015) Applying social media intelligence for predicting and identifying on-line radicalization and civil unrest oriented threats. CoRR. arXiv:1511.06858
13. Hale WC (2012) Extremism on the world wide web: a research review. Crim Justice Stud 25(4):343–356
14. Mei J, Frank R (2015) Sentiment crawling: extremist content collection through a sentiment analysis guided web-crawler. In: ASONAM. ACM, pp 1024–1027
15. Bouchard M, Joffres K, Frank V (2014) Preliminary analytical considerations in designing a terrorism and extremism online network extractor. In: Computational models of complex systems. Springer, pp 171–184
16. Chatfield AT, Reddick CG, Brajawidagda U (2015) Tweeting propaganda, radicalization and recruitment: Islamic state supporters multi-sided twitter networks. In: Dg.o. ACM, pp 239–249
17. Burnap P, Williams ML, Sloan L, Rana O, Housley W, Edwards A, Knight V, Procter R, Voss A (2014) Tweeting the terror: modelling the social media reaction to the woolwich terrorist attack. Soc Netw Anal Min 4(1):1–14
18. Zhou Y, Reid E, Qin J, Chen H, Lai G (2005) Us domestic extremist groups on the web: link and content analysis. IEEE intelligent systems 20(5):44–51
19. Buntain C, Golbeck J, Liu B, LaFree G (2016) Evaluating public response to the boston marathon bombing and other acts of terrorism through twitter. In: ICWSM. pp 555–558
20. Berger J, Strathearn B (2013) Who matters online: measuring influence evaluating content and countering violent extremism in online social networks
21. Wei Y, Singh L (2017) Using network flows to identify users sharing extremist content on social media. In: Pacific asian conference on knowledge discovery and data mining (PAKDD)
22. Rowe M, Saif H (2016) Mining pro-ISIS radicalisation signals from social media users. In: ICWSM pp 329–338
23. Cavazos-Rehg P, Krauss M, Grucza R, Bierut L (2014) Characterizing the followers and tweets of a marijuana-focused twitter handle. J Med Internet Res 16(6)
24. Bilge L, Strufe T, Balzarotti D, Kirda E (2009) All your contacts are belong to us: automated identity theft attacks on social networks. In: WWW
25. Lee S, Kim J (2012) Warningbird: detecting suspicious URLS in twitter stream. In: NDSS
26. Xu W, Zhang F, Zhu S (2010) Toward worm detection in online social networks. In: ACSAC. ACM, pp 11–20
27. Barbosa L, Feng J (2010) Robust sentiment detection on twitter from biased and noisy data. In: ACL pp 36–44
28. Jiang L, Yu M, Zhou M, Liu X, Zhao T (2011) Target-dependent twitter sentiment classification. In: HLT. ACL, pp 151–160
29. Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. LREc 10:1320–1326
30. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R (2011) Sentiment analysis of twitter data. ACL, pp 30–38
31. Li C, Zhang J, Sun J-T, Chen Z (2013) Sentiment topic model with decomposed prior. In: SIAM

32. Gutiérrez E, Cervantes O, Báez-López O, Sánchez JA (2015) Sentiment groups as features of a classification model using a spanish sentiment lexicon: a hybrid approach. In: Pattern recognition. Springer, pp 258–268
33. Saif H, He Y, Fernandez M, Alani H (2016) Contextual semantics for sentiment analysis of twitter. Inf Process Manag 52(1):5–19
34. Nasukawa T, Yi (2003) Sentiment analysis: capturing favorability using natural language processing. In: K-CAP. ACM, pp 70–77
35. Hu V, Liu B (2004) Mining and summarizing customer reviews. In: KDD. ACM, pp 168–177
36. Stanford corenlp sentiment analyzer. http://stanfordnlp.github.io/CoreNLP/sentiment.html
37. Vader sentiment analysis. https://github.com/cjhutto/vaderSentiment
38. Sentistrength. http://sentistrength.wlv.ac.uk/

Chapter 18
An Organizational Visualization Profiler Tool Based on Social Interactions

Panagiotis Karampelas

Abstract Complex organizational environments require highly-skilled employees who are both good at their everyday work and at the same time digitally literate, capable of using communication platforms and social media. Moreover, the familiarization of employees with technology and their tendency to bring their own devices at work, has created an additional headache for information security officers who fear that several backdoors can be opened to the organization security infrastructure not only by the misuse of the devices but also by a potentially highly-skilled employee. The proposed, in this chapter, social profiler tool aims at identifying potential inside threats using organizational information i.e., communication messages either from emails or social media. The information collected is then analyzed using a custom vocabulary which contains keywords related to the sensitive information of the organization in order to produce a list of employees who can potentially become insider threats. Finally, the social profiler tool incorporates six different visualizations of the employees under investigation with attributes such as their behavioral profile, ego network, word cloud, and a comparative profile of each employee in contrast to other profiles in their network. The tool's effectiveness has been tested with an actual business communication dataset using a well-established generic vocabulary demonstrating promising results.

Introduction

The contemporary organizational environment requires employees who are highly-trained not only in their line of business but also competent in using the diverse digital communication platforms utilized by their company such as email systems, intranets, webportals, social media platforms, enterprise systems, etc. Moreover, the tendency of allowing the employees to bring their own device (BYOD) at the working environment has become a headache not only to the administrators of the corporate

P. Karampelas (✉)
Hellenic Air Force Academy, Dekelia Air Base, Attica, Greece
e-mail: panagiotis.karampelas@hafa.haf.gr

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_18

systems but mostly to information security officers. In addition, the trend to move most of the corporate data and systems for easy access to mobile users and employees to the cloud adds another piece to the puzzle for the information security officers of the company. This complex business ecosystem leads 89% of companies to feel that they are vulnerable to insider attacks and 34% of them to feel ‘extremely vulnerable’ [19]. There are two types of insider threats; more specifically those that stem from the malicious employees and those that may arise due to an accidental insider [7]. Malicious insiders are those who will deliberately try to harm the enterprise either because they are employed by a competitor e.g., for espionage or even because they are discontented with the organization for any reason [22]. The second category of accidental insiders can become a danger to the company because of their negligence to the security policies of the organization, their poor training, their lack of interest and motivation or their laziness among other things [25]. In both cases, the insider threats are considered as a major concern in planning the information security policies of an organization [2] and a constant headache for information security officers and executive administration.

As a consequence, various bodies have attempted to develop a mitigation plan for insider threats. For example, there is the Intelligence National Security Alliance (INSA) Insider Threat roadmap [13], the Carnegie Mellon University’s Computer Emergency Response Team (CERT) insider threat best practices [21], and the Framework for Improving Critical Infrastructure Cybersecurity [16]. The mitigation plans in the majority of frameworks recommend specific actions such as access control, awareness and training, security in valuable assets, data, processes and guidance on incident management and recovery [16].

Apart from the governmental agencies, several security organizations and independent researchers have also worked on detecting an insider threat using different techniques. In the work by Balakrishnan [2], a validation framework is proposed for employees and contractors; various other technical solutions that monitor employees’ resources are proposed by Cole in 2015 [7]; psychological profiling is proposed by Nurse et al. in 2014 [17] and Brdiczka et al. in 2012 [4] while hybrid methods to detect insider threat are suggested by Eldardiry et al. in 2013 [9].

As it appears from the proposed solutions, surveillance and monitoring of employees is a technique that is proposed by the majority of the organizations and researchers. However, a key finding in a survey regarding the mitigation of insider threats was that 84% of companies do not scan the outgoing emails for confidential data that may be sent by employees [3] which is surprising since someone would expect that the first medium that a company would monitor in an attempt to protect their confidential data is communication channels.

Especially with the introduction of social networking platforms which immediately became popular, millions or billions of people started communicating through social media either to update their social ties about their current activities or to share personal thoughts. The initial negative attitude of the organizations against Facebook and other social network platforms regarding the productivity of their employees has changed towards positive exploitation of Facebook and social media in general for marketing purposes, helpdesk support, customer surveys and a ton of other applications that the company officials may invent. Recently, marketing campaigns that combine a number of digital communication channels have been developed increasing the impact of social media in business. On the other hand, social scientists with the assistance of computer science researchers have started developing methods and algorithms that could be used to study the new forms of social formations also called social networks. This is imperative since the traditional analysis methods fell short of revealing information about inherent characteristics of social networks, e.g., the structure or the origin of a clique in a social network, or the leader or the opinion influencer inside a clique [15]. The same applies when more complicated information is required regarding the social networks participants, e.g. their behavioral or political profile. Thus, everyday new methods are proposed under the field of social media analytics [11] or social network analysis and mining.

In this work, a new tool is proposed that utilizes a variety of analytical methods to provide insight in an organization regarding the communication profile and social online interactions of its employees. The tool takes advantage of existing social network analysis methods which are combined with newly developed visualization techniques to illustrate the profile of an individual. The prototype has been tested against the well-known email dataset from Enron [6] and classified the employees of Enron in specific categories based on a predefined dictionary [20]. Then the classification of the employees as well as their social interactions were illustrated with different visualization methods. The tool also dives into the content of their interactions revealing textual information which may be useful to the information security officers that are interested in discovering potential insider threats to their organizations. The prototype can also be customized to meet the requirements of any organization depending on the dictionary that will be used in order to monitor communications and classify the employees. It can be further extended to combine more than one communication channels and extract useful information from different social media platforms.

The rest of the chapter is structured as follows: section "Related Work" presents the related work in the area of insider threat detection and mitigation as well as some relevant social network analytics techniques. Section "Guiding Principles" describes the guiding principles of the development of the social profiler tool adopted in the current work. Section "Social Profiler Overview" presents the tool methodology and further discusses the analytics and visualization techniques used while the proof of concept is presented in section "Case Study" in the form of a case study with an example dataset. In the last section conclusions are drawn and future work is discussed.

Related Work

As the presented work is related with insider threat detection and social network analysis, the specific section has been divided in two parts that present the relevant work in both areas.

Insider Threat Mitigation Techniques

As the proposed tool aims to assist information security officers to acquire a better view of the employees' behavior, it is interesting to also review the monitoring methods that have been proposed to mitigate insider's threat. The methods are usually classified into technical monitoring solutions and those examining psychological, personal or other characteristics of the employees. In the following paragraphs, some of the more prevalent techniques in these categories will be presented.

In the work of Cole at 2015 [7], a series of technical solutions are presented to mitigate the insider threats. These solutions vary from using inbound and outbound proxies with content blocking and filtering, web filtering, sandboxing of the incoming executables, application whitelisting, data classification and netflow analysis to detect data exfiltration, user activity monitoring and Security Information and Event Management (SIEM) systems that allows the companies to centrally collect all the alerts from network and hardware systems. All the aforementioned technical solutions are indeed useful and collect a wide range of information that will potentially discover a threat from inside if it is appropriately analyzed by specialized personnel. However, if the insider is a highly-skilled and technically competent employee who has elevated access then it is very easy for him or her to overcome the limitations or to hide the tracks of a malicious activity that took place. Another issue with the specific methodological approach is that the setup of such an infrastructure is rather complicated and requires significant investment in the infrastructure and administration of such systems.

Balakrishnan from SANS, in his work [2] also proposes a validation framework for employees and contractors in order to monitor their activities and raise the necessary alerts in case of suspicious behavior. The proposed framework sets a number of indicators that could potentially indicate suspicious activity inside the network of a company. These indicators comprise data uploading or downloading activity, use of non-approved applications and tools, access to blacklisted websites, multiple attempts to connect to resources without authorization, use of confidential keywords or data, excessive use of network, etc. While the specific measures can indeed provide an indication of a suspicious activity, the setup and the implementation of such a framework requires a variety of tools which are rather complicated and require advanced technical skills for the company administrators to take them to their full potential.

Other proposals include less technologically advanced solutions which are based on surveillance. For example, Spitzner proposes the creation of honeypots or 'honeytokens' inside the resources of an organization with tempting titles e.g. strategic plan of the company [22]. The honeytokens can be items such as documents or emails that belong to high rank officers of the company e.g., Chief Executive Officer, Vice President, etc. who will be informed about the existence of such a honeytoken and they will never access it. If someone else accesses it, then it is an indication that the specific employee may be an insider. Eldardiry et al. [9] propose the constant monitoring of the employees' activity and detection of behavioral inconsistency. The domain of activities that are monitored for anomalies are the logon behavior, the device used, file access behavior, internet usage, and email activity. The activity logs are then analyzed using three models to detect the anomalies in the behavior of the employee.

Another approach that has been adopted by several researchers in order to mitigate the insider threat is that of the proactive personality or behavioral analysis of the employees. In this context, Kandias et al. presented a model that collects user characteristics by a "Psychological Profiling component" and user usage information from the various information systems. Then the information collected is combined and analyzed by a component called "Decision Manager" which finally concludes whether the specific user constitutes an insider threat or not [14]. According to the authors, the user characteristics are determined applying the Social Learning Theory by examining three aspects of the personal profile of the employee, namely his/her sophistication, his/her predisposition to malicious behavior, and his/her stress level. While the results were the desirable ones according to the authors, the dataset in which the model was applied was based on YouTube comments and not in a real organization environment.

Brdiczka et al. presented a model that is based on Structural Anomaly Detection using information from social networks and Psychological Profiling based on behavioral patterns [4]. More specifically, the authors attempt to detect communications that are beyond the expected ones with peers or supervisors e.g., with outsiders, or to find topics in the communication of the employees that are beyond the typical ones in a business environment. Moreover, an attempt is presented to classify employees into personality models monitoring their traits, their emotional state, etc. and to monitor sudden changes in any of the aforementioned attributes. The specific model was applied in a very large number of gamers who participated in the World of Warcraft (WoW) game and it was proved that the model could work to predict anomalous behavior, however the context of the experiment was not relevant to a real business environment.

Nurse et al. presented a framework that conceptualizes the different aspects of an insider attack taking into consideration the main actor characteristics as well as the attack characteristics [17]. The specific framework is based on the observation of several employee characteristics such as physical behavior, historical behavior, cyberspace behavior and also on the assessment of various other characteristics such as psychological state, motivation for an attack, given opportunities to attack, technological skills, role in the organization, etc. The framework was validated in three real cases and the authors managed to capture the key elements of the attack in each case based on the framework analysis. However, there is no concrete way of applying the framework in a real business environment and automating the collection of the artifacts required and then assessing the employees against these characteristics before an attack has taken place.

Social Networks Analytics Techniques

Social networks analysis has become a very popular means to analyze the interaction between individuals either in social media, email communication, collaboration or any other activity between individuals that can be recorded electronically. As such, social networks analysis can play an important role in the attempt to proactively detect insider threat. Particularly email communication as it was presented in the previous section plays an important role in the detection of insider threats and thus there are already various methods that allow researchers to analyze email communications for various purposes.

The email analysis techniques have been used for spam classification, as in the case of the Email Mining Toolkit (EMT) and Profiling Email Toolkit (PET) by Hershkop and Stolfo [12]. The specific tool can analyze the emails of an organization using machine learning algorithms in order to detect spam emails or virus propagation [23]. The PET framework is the corresponding real-time version of EMT enriched with additional email analysis algorithms which improve the accuracy and efficiency of the framework. However, both tools do not explore or visualize the social dynamics of the correspondents and their social network.

Other researchers have worked on detecting discussions in the email communication [1] inspecting various characteristics of the communication such as the terms used, the individuals communicating and the date and time of the communication. No further analysis is taking place as for example to identify suspicious behavior or to further understand the attitude of the members of the discussion.

The email analysis has also been applied in digital forensics analysis [8] to classify behaviors or persons into professional roles. Another application of email analysis in the same work was authorship, namely attribution of a text to an author with a success rate up to 70%. As in the previous cases, email analysis was not used proactively to detect suspicious behaviors or possible insider threats.

A work more similar to the objective of the proposed tool can be found in Persaud and Yong' s framework [18] which extracts structured and unstructured information from emails and using feature detection attempts to identify hidden relationships between the correspondents e.g. if they use a specific set of predefined words. The specific framework, though, does not provide any additional tools to further explore suspicious relations.

Van Alstyne et al. have proposed a social network analysis tool that analyzes email communication and visualizes various attributes of the communication network [24] which potentially could be used to further study a suspicious relationship between employees who may constitute an insider threat. The tool, however, does not expand to other tokens which may be interesting in case of a potential insider threat.

Finally, the idea of email analysis in the tool presented in this chapter was partially inspired by the work of Zhou in [27, 28], however with a totally different objective. The purpose in the work of Zhou was to investigate whether people who correspond together share the same ethical values using the set of universal values of Schwartz [20]. To better prove the hypothesis, Zhou further analyzed the email correspondence applying various social networks mining techniques such as degree distribution analysis, centrality analysis and overlapping cluster analysis. It is worth mentioning, that the dataset most commonly used to test the email analysis techniques is that of Enron [6] as it is the only one publicly available for research purposes and is very relevant to the organization culture in a business environment in which the proposed tool operates.

Guiding Principles

Building upon the previously presented related research, the proposed tool aims to assist the information security officer to detect potential insider threats by surveilling the social interactions of the employees. More specifically, the tool falls under the category of social network surveillance tools and explores the behavior and psychological characteristics of the correspondents by analyzing various features collected from diverse sources of social interactions i.e. emails, twitter, Facebook, etc. The collected information is then visualized using multiple visualization methods in order to provide the analyst with insight of a potential insider threat to the organization. The visualization of the social network and social attributes of the employee can lead the analyst to better understand the social environment of the employee and potential threats to other employees. The tool could also be used as an early warning system for extreme or radicalized views of employees inside a sensitive environment, such as a company or institution operating in the field of national security or defense that requires loyalty and confidentiality on behalf of its employees. Overall, the tool fits the category of the proactive tools for detecting insider threats using social network analytics. Social network analytics, according to the CUP framework [11], require three operational phases: “capture, understand and present”.

Data capturing: Capturing data in general is not such a simple process since quite often there are several technical and ethical issues. Technical issues depend on the type and number of data sources. If the data come from different sources then they most probably contain either unlinked or overlapping information. In both cases it is a problem, since in the first case the unlinked information needs to be associated or normalized depending on the case in order to be useable, otherwise it will be fragmented without any practical value; in the second case, the overlapping information needs to be detected and separated in order to avoid skew results. An additional problem when capturing data from a source is how clean the dataset is which means whether there is noise or missing data in the dataset and as a result the analysis cannot be completed or may not provide accurate results. Noise or missing data may occur in case there is no direct access to the data source or if the data capturing is performed indirectly. Ethical issues come to light whenever a person is involved at any stage of the production of the data and thus there is an indication that the data may contain private information about the person. In a social network, both issues arise since most of the times the communication is not formal and may contain partial sentences or acronyms or attachments combined with personal information about the user.

The setting for the recommended use of the proposed tool is a business environment where both the data sources and the data are more structured. For example, the communications are usually performed with predefined tools that are easily to be monitored since the majority of them are hosted within the organization or the organization has administrative rights when the infrastructure is cloud-based. The employees are obliged to use the same tools for consistency and they usually follow specific guidelines or rules e.g., signature in the case of emails or specific hashtags in the case of twitter, etc. In addition, the account names provided by the organization follow the same pattern or rules and in most cases, there is a unique username that identifies the employee in all the different business systems. Thus, the social interactions performed within a controlled environment usually do not suffer from unclean or missing data and are easily normalized since there is a common identifier. The ethical issues in a business environment similarly can be easily overcome since the resources and the infrastructure used are provided by the company and thus the company can claim that the infrastructure is provided for the productivity of the employees and not for personal communications etc. and thus the company retains the right to monitor the resources and infrastructure for appropriate use. Of course, the organization needs to abide by the local legal and ethical rules and ensure that the surveillance process is performed for the protection of the company and the employees and cannot be exploited either by the management of the organization or other employees for any reason. The employees at the same time need to be aware of how the surveillance system works and how the information recorded is analyzed and used. In most countries and companies this is achieved through the orientation of the employee to the new job and the acquisition of the written consent that the employee understands and allows the organization to record information related to the everyday business.

Data understanding: The next phase in social network analytics concerns the understanding of the data collected in the data capturing. There are several techniques that can be applied for analyzing the data collected which depend mainly on the actual topic of interest in each case. When it comes to the analysis of social networks, the techniques proposed usually depend on the social attributes that will be analyzed. For example, if the focus is on the important persons in a social network then the respective techniques are utilized to calculate e.g., the closeness or betweenness centrality or any other social network measure. When the focus is on the text written by an individual then the corresponding text mining and analytics techniques need to be utilized. In the case of the proposed tool, since a multilevel analysis is performed, several different analytics methods mostly from the area of text mining and social networks analysis are employed as it will be further discussed in the following sections.

Data presentation: Data representation is equally important in social network analytics since the datasets used usually include some hundreds or thousands of data points that are not easy to visualize. For that purpose, several techniques have been developed in order to portray the most interesting attributes of the dataset depending on the case. In the proposed tool, some novel visualization techniques have been implemented that will assist the analyst to acquire a better insight of the behavioral and psychological profile of the employees and easily track their social interactions. The specific visualization techniques will be presented in the corresponding sections.

Social Profiler Overview

The profiler aims at identifying potential insider threats in an organizational environment by monitoring the content of the social interactions of the employees and their corresponding social network. This is achieved through the text analysis of the exchanged messages between the employees and/or external actors either in the form of email messages or tweets. By analyzing the produced text by any actor against a predefined dictionary with bags of words, it is possible to classify employees into different categories that correspond to the words used and in that way, discover potential malevolent employees. At the same time, social network analysis can assist in discovering cliques that are either supporters of the malevolent employee or potential targets of the employee based on some social networks metrics that can be extracted from the dataset with the collected information. The tool comprises all the necessary visualization techniques to assist the analyst to reach a conclusion based on the analysis performed. The proposed tool has been tested with a known dataset [6] and a predefined dictionary [20]. During the testing the proposed tool classified correspondents in different categories. To achieve the aforementioned objectives, the profiler engages six different visualizations. The first three visualizations explore the social profile of the actor under investigation from a different perspective while the rest of the visualizations illustrate the behavioral profile of the actor based on the textual analysis of the communication record of the actor (Fig. 18.1).

Social Profiler Methodology

The overall profiler methodology is illustrated in Fig. 18.2 where the different processing stages are presented. At the first stage, the tool agglomerates the relevant data from the different data sources. In this stage, the communication traffic produced by the employees is captured and stored in the corresponding infrastructure for further processing. In the current version, the tool monitors email traffic and twitter posts and captures all the ingoing and outgoing email messages and tweets. Each message or tweet is then analyzed and various attributes are extracted. The attributes that are essential for analyzing an email message are the subject of the message, the body of the message, the sender of the message, the recipient(s) of the message, the type of the recipient (to, cc or bcc), the time stamp of the message, and the file names of

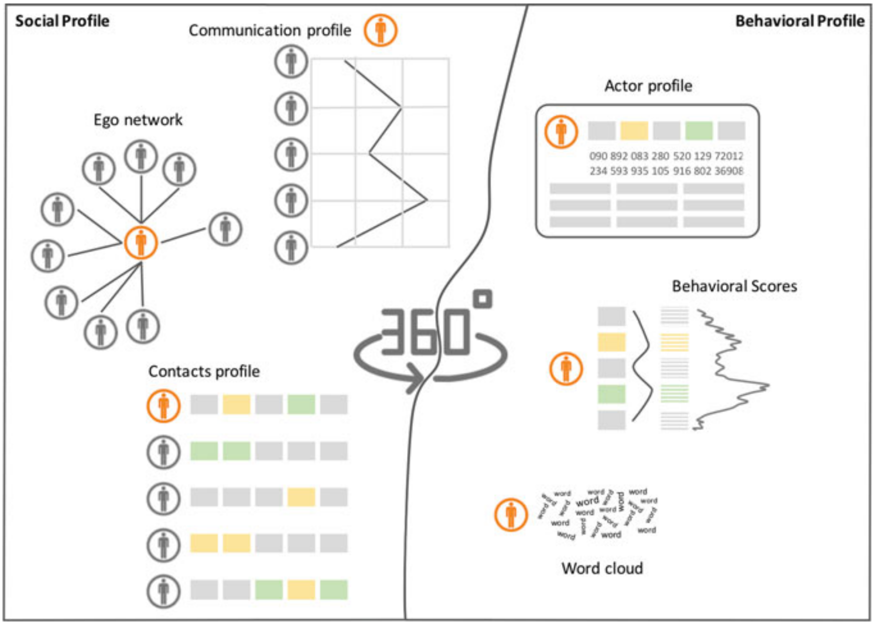


Fig. 18.1 Social profiler overview

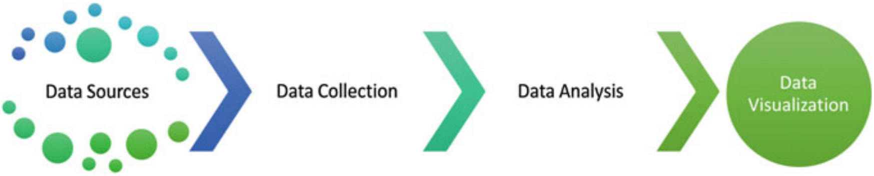


Fig. 18.2 Social profiler workflow

any attachments. In tweets, the screen name is captured, the text and the number of retweets and the screen names of the accounts who have retweeted the original text. Thus, let' s define as $C_i = \{c_{i1}, c_{i2}, \ldots, c_{in}\}$, the communication record of the actor i who has sent n messages or tweets and c_{ij} the jth the content of the message found in the record of the actor i.

For each recorded message sent by the actor, an additional vector that contains the required meta-information of the message is stored. The vector is defined as $M_{ij} = \{m_{ij1}, m_{ij2}, \ldots, m_{ijk}\}$, where m_{ijk} is a different meta-information sent by the actor i in the jth message. The different m_{ijk} records the type of the message i.e., email or tweet, the receivers, the timestamp, the geolocation, number of retweets, etc. depending on the availability of the meta-information. In case there is a message from a new source that appears for the first time then a new entry is created in the vector C and the meta-information of the message is respectively recorded in M vector. All subsequent messages from that specific actor are then appended to the specific record.

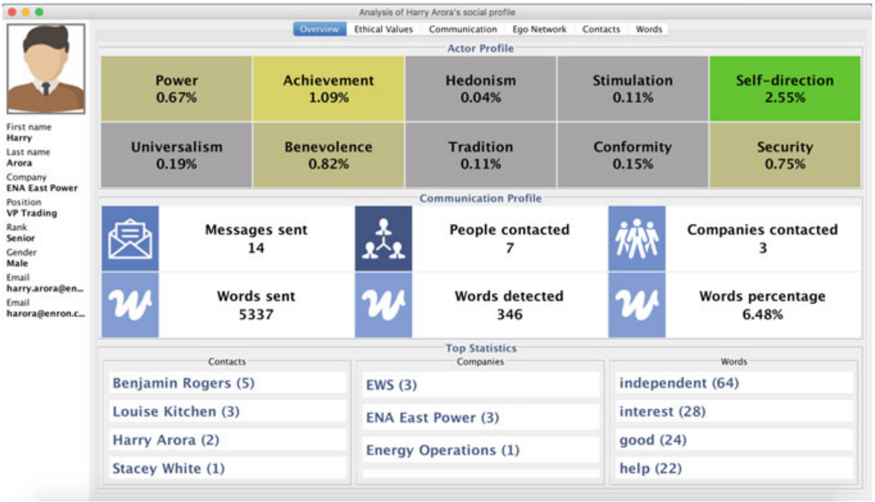
The next stage entails the cleaning of the message from any irrelevant information or information that can skew the results of the analysis. Such information is the duplicated content as for example when there is a response in an email message that contains the original message. The system eliminates the duplicated information maintaining only the new message. In addition, a number of 668 English stop words are removed from the message such as single characters, special characters, punctuation symbols, articles, prepositions, conjunctions, and some basic words that are not taken into consideration. The analyst though can customize the specific list of stop words that are removed according to the desired objectives. No other preprocessing takes place since the profiler needs to be sensitive and distinguish words that may, for example, have the same stem but can be used differently by the actors. An example, of how words with the same stem are handled is the derivatives of the word 'power'. Depending on the dictionary of the words of interest, the analyst may desire to monitor the words 'power', 'powerful', 'powerless', 'powered', 'overpower', 'empower', etc. and not to identify a generic reference to the word 'power'. The result of the preprocessing phase is a new vector for the communication record of the actor as follows: $C'_i = \{C'_{i1}, C'_{i2}, \ldots, C'_{in}\}$ where each $C'_{ij}, j = 1, 2, \ldots, n$ contains the important words found in the jth message of actor i. The next step is the actual data analysis of the communication record of a specific actor. An actor can be selected and then a number of different data analytics tasks take place to produce six different visualizations.

Social Profiler Visualizations

The first visualization presents the overview of the selected actor's profile displaying a number of statistics and the ethical profile as it has been created based on the collected communication records. More specifically, the metrics that are displayed in the specific visualization are the ethical profile, the communication profile and the top statistics of the actor's communication (Fig. 18.3).

Actor Profile

The actor profile is calculated based on a dictionary of predefined words that have been set by the company as indicative of suspicious communication. The vocabulary should be organized in categories and subcategories each of which is associated with a number of related words. The profiler can then use the specific vocabulary and based on the analysis of the communication can classify the actor into the respective category—subcategory scheme. Thus, the vocabulary can be defined as $V = \{V_{11}, V_{12}, \ldots, V_{nm}\}$ where V_{kl} is the k category and l subcategory of the vocabulary with $V_{kl} = \{w_1, w_2, \ldots, w_s\}$, $k = 1, 2, \ldots, n$, $l = 1, 2, \ldots, m$ and w_t it the tth word of the vocabulary category k and subcategory l. The classification algorithm



First name
Harry
Last name
Arora
Company
ENA East Power
Position
VP Trading
Rank
Senior
Gender
Male
Email
harry.arora@en...
Email
harora@enron.c...

Actor Profile

Power 0.67%	Achievement 1.09%	Hedonism 0.04%	Stimulation 0.11%	Self-direction 2.55%
Universalism 0.19%	Benevolence 0.82%	Tradition 0.11%	Conformity 0.15%	Security 0.75%

Communication Profile

Messages sent 14	People contacted 7	Companies contacted 3
Words sent 5337	Words detected 346	Words percentage 6.48%

Top Statistics

Contacts	Companies	Words
Benjamin Rogers (5)	EWS (3)	independent (64)
Louise Kitchen (3)	ENA East Power (3)	interest (28)
Harry Arora (2)	Energy Operations (1)	good (24)
Stacey White (1)		help (22)

Fig. 18.3 Overview of the employee profile

retrieves all messages of actor i and tests whether a word $C'_{ij} \in \{c^i \mid \Phi_r(c')\}$ where $\Phi_r(c') = \begin{cases} 1, & c' \in V_r \\ 0, & c' \notin V_r \end{cases}$. The output of this subtask is a vector R_i that contains in each position the number of occurrences for each word belonging to V_{kl} for $k = 1, 2, \ldots, n$ and $l = 1, 2, \ldots, m$ for the actor i. The algorithm then transforms the vector $R_i \xrightarrow{f(r)} R'_i$ where $f(r) = \sum_{t=1}^{s} w_t$ that contains the score for each category and subcategory. Then the new vector R'_i is normalized to contain in each position the normalized value for each category and subcategory based on the total number of words used from the specific actor. The final classification of the employee is achieved using a set of rules applied to the normalized vector $\hat{R}'_i$. Initially, the actor is assigned to the category where his/her score is the highest and then is recursively assigned to other categories based on the weighted Euclidean distance of the categories. The weights can be set according to the value of each category of the vocabulary that needs to be tested against the communication messages.

Communication Profile

The specific metrics visualize the key communication factors of the actor. More specifically, the total number of communication records namely the size of the vector C'_i is displayed, the number of the recipients of the messages which is retrieved by the meta-information vector M_{ij}, the number of different companies contacted which is derived by the known actors contacted, the total number of words sent which is the $sum_{t=1}^{s} w_t$ where $w_t \in C'_{ij}$ which is the jth message of actor i, the number of the occurrences of the dictionary words which is calculated as the $sum_{k=1}^{m} r_k$ where $r_k \in R_i$ is the number of the occurrences of the words that belong to the vocabulary in the messages and the percentage of the vocabulary words per total words in the messages. The communication profile metrics provide context to the calculated values in the actor profile since the analyst can see e.g., the overall messages sent by the actor and understand whether the use of the specific words of the dictionary prevails or if the designated words are just an insignificant minority.

Top Statistics

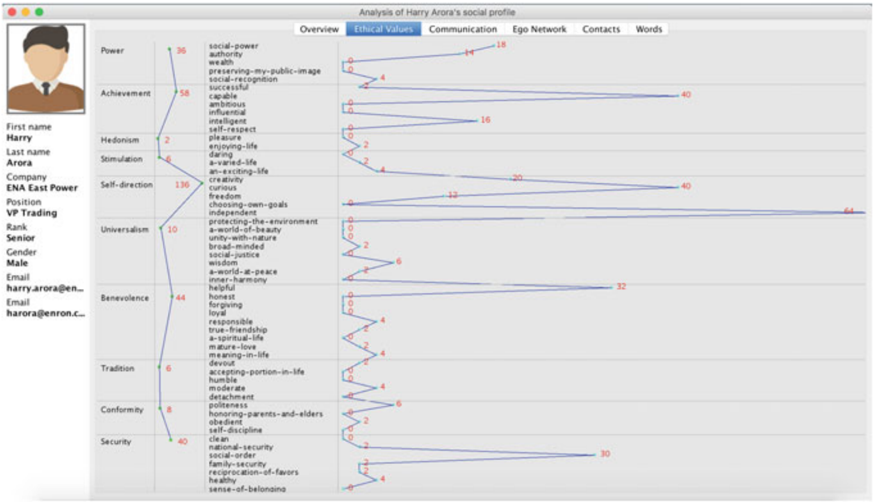
Top statistics provide insights to the analyst regarding the most frequent communication artifacts of the actor. More specifically, the top four people who communicate more frequently with the actor under investigation are presented, which is calculated using the meta-information vector M_{ij}, the top four companies contacted which is also derived by the other actors contacted and the top four words used that are selected as the four top occurrences found in vector R_i of the vocabulary words. In each entry presented in the list, the actual number of occurrences is presented in parentheses in order to assist the analyst to better understand the actual scores in each category.

Behavioral Profile

The ethical values visualization (Fig. 18.4) presents the overall score of the actor against the categories and subcategories of the vocabulary as it was presented in [26]. The visualization derives its values from vector R_i in order to display the full details of the analysis. The visualization chart that is used is a composite line plot diagram that illustrates the values each category and subcategory has scored. The left part of the chart displays the category scores with the total number of the words found for each category and the right part of the chart displays the subcategories per category and the corresponding number of occurrences of the words in each subcategory. The total number of words on the left side of the diagram should agree with the sum of the words on the right side of the diagram per category. To improve the visualization of all the scores, the composite chart adjusts the viewpoint on both parts of the diagram to include all the values.

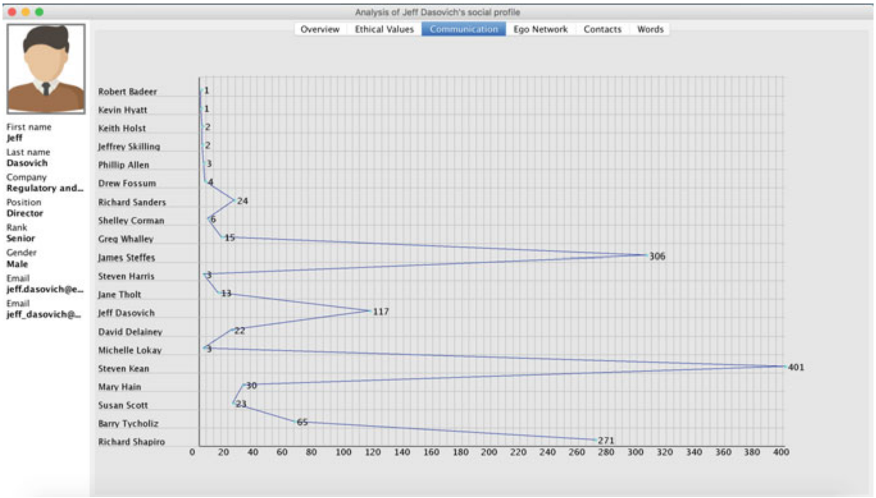
Actor Communication

The next visualization available to the profiler displays a frequency distribution diagram of the communication with the other actors. The number of the communication messages exchanged is calculated based on the meta-information vector M_{ij} where j is the meta-information vector of the jth message of the ith actor. From this vector, the recipients' addresses are counted and then the resulting vector is illustrated in


Analysis of Harry Arora's social profile

First name
Harry
Last name
Arora
Company
ENA East Power
Position
VP Trading
Rank
Senior
Gender
Male
Email
harry.arora@en...
Email
harora@enron.c...

Fig. 18.4 Ethical values visualization [26]


Analysis of Jeff Dasovich's social profile

First name
Jeff
Last name
Dasovich
Company
Regulatory and...
Position
Director
Rank
Senior
Gender
Male
Email
jeff.dasovich@e...
Email
jeff_dasovich@...

Fig. 18.5 Visualization of actor's communication

the diagram (Fig. 18.5). It is noteworthy that in case of email messages the system also counts those emails that have been sent by the actor to himself/herself which is an indication for the importance of a message.

Ego Network

A very important visualization for the profiler is the ego network visualization [26] which integrates several different attributes of meta-information in one diagram. Ego-networks are created for a specific individual and include all the ties with the actors the individual interacts with [10]. The advantage of this type of visualization is that it facilitates the study of the micro world of the actor and infers "statistically significant" conclusions for the whole network instead of studying the whole social network with all the actors which is a very complex graph. The diagram can also be enriched with supplementary information regarding the actors or the ties and thus provide a better view of the social network. The specific visualization as it is described in [26] is based on a radial layout which presents the node (actor) under investigation and all the ties of this specific node in consecutive arcs at the center of the graph. The closer to the center a node is, the denser is the communication of the central node with this node. This is also illustrated in the diagram by thicker lines which represent the ties. In each node, the number of the exchanged messages is also illustrated. An alternative visualization of the most frequently contacted nodes can be displayed by using different size for the node or gradient colors. To better depict the difference between the actors with frequent communication against those with sparse communication the diagram is enhanced by presenting a number of concentric circles. The number of the concentric circles is calculated based on the number of the intermediate nodes between the node with the minimum communication and that with the maximum number of received messages (Algorithm 1). The layout algorithm (Algorithm 2) of the specific diagram introduces a different approach from the standard ego network regarding the positioning of the nodes in the diagram. The algorithm creates clusters of actors based on their meta-information by grouping the actors who belong to the same entity together. In that way, actors who belong to the same company or department are placed together in the same sector and the name of the department or the company is visualized at the top of the sector. The area of each sector represents how dense the communication with the individuals of the specific company or department is. If the sector area is large, it means that the central actor has an intense communication with representatives of the specific company or department, otherwise the area will be small. Each sector is defined by two radii whose central angle is calculated based on the percentage of the messages sent to actors of a specific company or department (Algorithm 2).

Algorithm 1 Drawing of the Concentric Circles [26]

```
Input: Adjacency list $R_l$
$c \leftarrow width \,/\, 2$
$color \leftarrow [blue, white]$
for $i = 0$ to $R_l.value.size - 1$ do
    $r_{min} \leftarrow c \,/\, \ln(1 + \min(R_l[i].value))$
    $r_{max} \leftarrow c \,/\, \ln(1 + \max(R_l[i].value))$
end for
$n \leftarrow \ln(|r_{max} - r_{min}|)$
$r_{pr} \leftarrow 0$
for $q = 1$ to $n$ do
    $r_q \leftarrow q * r_{min}$
    $drawcircle(width \,/\, 2 - 10, height \,/\, 2 - 10, r_q)$
    $fillsector(r_{pr}, r_q, color[q \mod 2])$
    $r_{pr} \leftarrow r_q$
end for
return
```

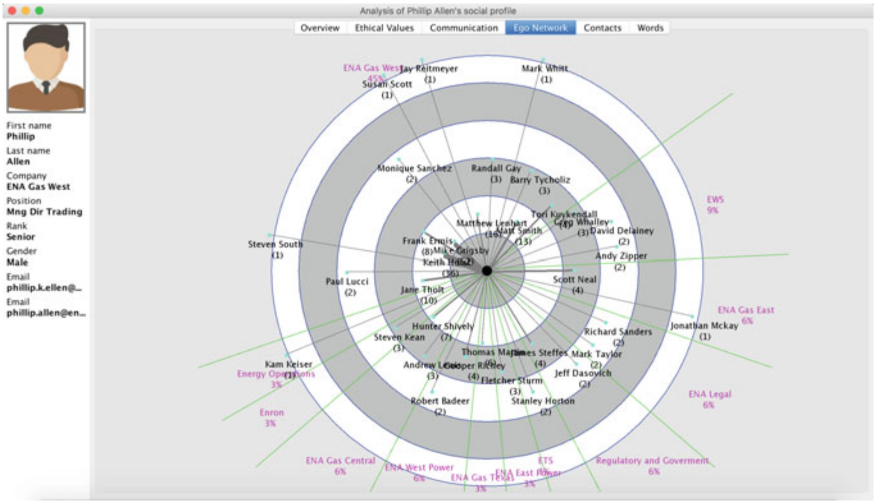
Algorithm 2 Drawing of the Department Sectors [26]

```
Input: Adjacency list $R_l$
$\phi \leftarrow 0$
$\theta \leftarrow 0$
$c \leftarrow width \,/\, 2$
for $i = 0$ to $R_l.value.size - 1$ do
    $r \leftarrow c \,/\, \ln(1 + \min(R_l[i].value))$
end for
for $q = 0$ to $max(R_l.key) - 1$ do
    $\phi_t \leftarrow \phi$
    $\theta_t \leftarrow \theta$
    $\theta \leftarrow \theta + (R_l[q].key * 2 * \pi \,/\, sum(R_l(q)))$
    $x \leftarrow (width \,/\, 2) - 10 + (r + 100) * \sin\theta$
    $y \leftarrow (height \,/\, 2) - 10 + (r + 100) * \cos\theta$
    if $(R_l.size \# 1)$ then
        $draw((width \,/\, 2) - 10, (height \,/\, 2) - 10, x, y)$
    end if
    $\phi \leftarrow \phi_t \,/\, 2 + \theta \,/\, 2$
    $x \leftarrow (width \,/\, 2) - 10 + r * \sin\phi$
    $y \leftarrow (height \,/\, 2) - 10 + r * \cos\phi$
    $\theta_t \leftarrow \theta_t - \theta$
    $write(R_l[q].key.name, -100 * (\theta_t \,/\, 2 * \pi), x, y)$
end for
return
```

Algorithm 3 draws the nodes of the actors around the central node based on the previously mentioned rules. Thus, the nodes are sorted based on the entity they belong to and then they are drawn based on their distance from the center of the diagram. Fig. 18.6 illustrates the ego network as it is displayed by the profiler.

Algorithm 3 Drawing of the Nodes [26]

```
Input: Adjacency list $R_l$
$c \leftarrow width / 4$
$w \leftarrow 10$
$h \leftarrow 10$
$color \leftarrow cyan$
$n \leftarrow R_l.value.size$
for $i = 0$ to $n - 1$ do
    $r \leftarrow c / \ln(1 + R_l[q].value)$
    $x \leftarrow (width / 2) - 15 + r * \sin(2 * \pi / n * q)$
    $y \leftarrow (height / 2) - 15 + r * \cos(2 * \pi / n * q)$
    $drawmode(x, y, w, h, color)$
    $write(R_l.value.name, R_l.value.count)$
end for
return
```

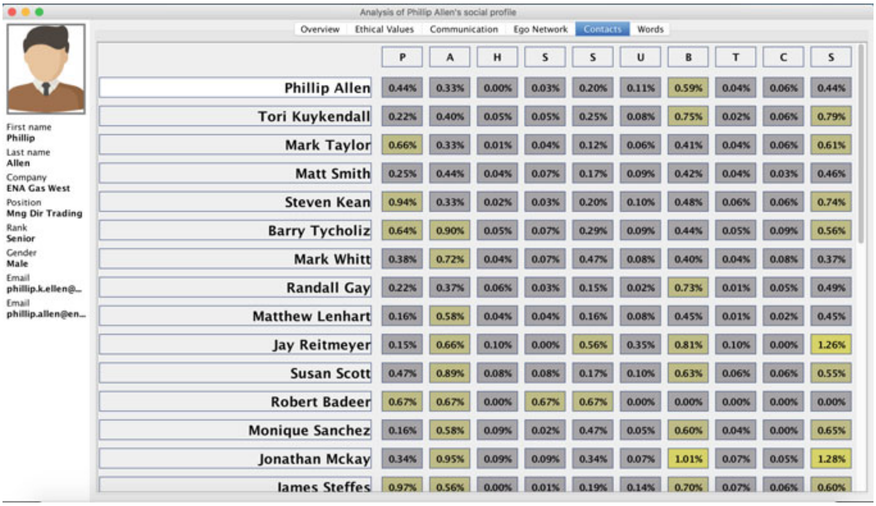


First name
Phillip
Last name
Allen
Company
ENA Gas West
Position
Mng Dir Trading
Rank
Senior
Gender
Male
Email
phillip.k.ellen@...
Email
phillip.allen@en...

Fig. 18.6 Ego network of an actor [26]

Contacts Profile

Another useful visualization which is available in the profiler is the comparative presentation of all actors' scores in the ten different categories. The visualization resembles the adjacency matrix of nodes however instead of presenting the number of messages exchanged between the actors, the calculated score of each actor against the predefined categories is presented. The profile of the actor under investigation is presented in the first line of the matrix and then all the profiles of the actors who communicate with the main actor are listed below (Fig. 18.7). The column headers represent the different categories defined by the analyst while the cell values below correspond to the calculated score of each actor. The score as mentioned above is

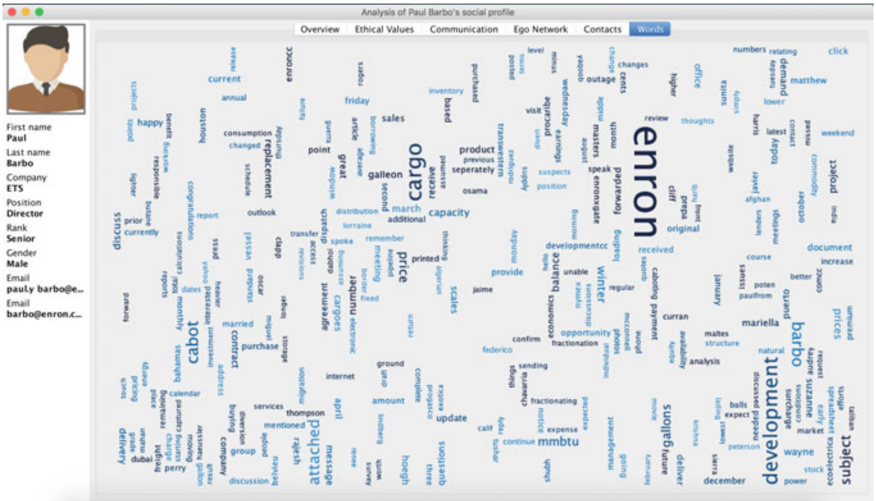


First name
Phillip
Last name
Allen
Company
ENA Gas West
Position
Mng Dir Trading
Rank
Senior
Gender
Male
Email
phillip.k.ellen@...
Email
phillip.allen@en...

	P	A	H	S	S	U	B	T	C	S
Phillip Allen	0.44%	0.33%	0.00%	0.03%	0.20%	0.11%	0.59%	0.04%	0.06%	0.44%
Tori Kuykendall	0.22%	0.40%	0.05%	0.05%	0.25%	0.08%	0.75%	0.02%	0.06%	0.79%
Mark Taylor	0.66%	0.33%	0.01%	0.04%	0.12%	0.06%	0.41%	0.04%	0.06%	0.61%
Matt Smith	0.25%	0.44%	0.04%	0.07%	0.17%	0.09%	0.42%	0.04%	0.03%	0.46%
Steven Kean	0.94%	0.33%	0.02%	0.03%	0.20%	0.10%	0.48%	0.06%	0.06%	0.74%
Barry Tycholiz	0.64%	0.90%	0.05%	0.07%	0.29%	0.09%	0.44%	0.05%	0.09%	0.56%
Mark Whitt	0.38%	0.72%	0.04%	0.07%	0.47%	0.08%	0.40%	0.04%	0.08%	0.37%
Randall Gay	0.22%	0.37%	0.06%	0.03%	0.15%	0.02%	0.73%	0.01%	0.05%	0.49%
Matthew Lenhart	0.16%	0.58%	0.04%	0.04%	0.16%	0.08%	0.45%	0.01%	0.02%	0.45%
Jay Reitmeyer	0.15%	0.66%	0.10%	0.00%	0.56%	0.35%	0.81%	0.10%	0.00%	1.26%
Susan Scott	0.47%	0.89%	0.08%	0.08%	0.17%	0.10%	0.63%	0.06%	0.06%	0.55%
Robert Badeer	0.67%	0.67%	0.00%	0.67%	0.67%	0.00%	0.00%	0.00%	0.00%	0.00%
Monique Sanchez	0.16%	0.58%	0.09%	0.02%	0.47%	0.05%	0.60%	0.04%	0.00%	0.65%
Jonathan Mckay	0.34%	0.95%	0.09%	0.09%	0.34%	0.07%	1.01%	0.07%	0.05%	1.28%
James Steffes	0.97%	0.56%	0.00%	0.01%	0.19%	0.14%	0.70%	0.07%	0.06%	0.60%

Fig. 18.7 The profile of the actor's contacts

calculated based on the frequency of the vocabulary words found in the communication messages. The use of color to display the different scores in each category indicates how relevant to the specific category the actor's scores are. The threshold for each color value can be defined by the analyst in order to adjust the warning levels according to the severity of each category. Moreover, the contacts matrix is interactive and the analyst can select another actor and review his/her corresponding profile.



First name
Paul
Last name
Barbo
Company
ETS
Position
Director
Rank
Senior
Gender
Male
Email
paul.y barbo@e...
Email
barbo@enron.c...

Fig. 18.8 The word cloud of the actor's communication messages

Word Cloud

The last visualization supported by the profiler is the word cloud of all the messages sent by the actor under investigation. This information can be used by the analyst to understand what the dominant words used by the actor are and form a better overview not only of the keywords defined by the analyst but of the actor's whole vocabulary (Fig. 18.8). The word cloud is constructed based on the vector $C_i = \{C_{i1}, C_{i2}, \ldots, C_{in}\}$ where i is the corresponding actor and c_{ij} is the corresponding jth message sent by the actor using the kumo java word cloud library [5].

Case Study

Description

For evaluating the performance of the Social Profiler, a case study was planned and carried out. The actual question that needed to be answered was how the social profiler behaves in a real organizational setting with hundreds of employees and hundreds of messages exchanged. For that purpose, a simulation of such an environment was set using a dataset with communication messages that is freely available and comes from a business environment (Enron). The dataset was initially published in May 2002 by Federal Energy Regulatory Commission (FERC) after the Enron scandal [6]. The initial dataset had several integrity issues which were progressively addressed by various researchers, such as Melinda Gervasio [6]. The most updated version that was used in the case study was published on May 7, 2015 and contains approximately 500,000 email messages from 149 employees. The time span of the exchanged messages starts in the late 1998 to mid 2002.

The next step in configuring the Social Profile is the selection of the appropriate vocabulary which could detect people in the company who can potentially become an insider threat. With that in mind, it was decided to work with a known vocabulary which has already been used by other researchers in order to be able to compare the actual results of the analysis and test the different visualization techniques. Thus, the vocabulary that was selected was the universal values set by Schwartz [20] which had already been used by Zhou [27] in the automatic text analysis of Enron email dataset. Zhou used a vocabulary of 663 words distributed in 10 categories and 54 subcategories as it is illustrated in Tables 18.1 and 18.2.

Actors' Analysis

Once the vocabulary has been set, the analysis of the dataset can start. The analyst can configure the sensitivity of the alert mechanism by defining the relevant thresholds

Table 18.1 Categories and subcategories of value words [20]

#	Category	Subcategory	Number of words
1	Achievement	Ambitious	7
2		Capable	10
3		Influential	4
4		Intelligent	23
5		Self-respect	9
6		Successful	5
7	Benevolence	A-spiritual-life	5
8		Forgiving	13
9		Helpful	15
10		Honest	23
11		Loyal	14
12		Mature-love	11
13		Meaning-in-life	8
14		Responsible	11
15		True-friendship	19
16	Conformity	Honoring-parents-and-elders	1
17		Obedient	27
18		Politeness	15
19		Self-discipline	8
20	Hedonism	Enjoying-life	14
21		Pleasure	8
22	Power	Authority	20
23		Preserving-my-public-image	10
24		Social-power	7
25		Social-recognition	3
26		Wealth	16
27	Security	Clean	21
28		Family-security	23
29		Healthy	16
30		National-security	20
31		Reciprocation-of-favors	6
32		Sense-of-belonging	8
33		Social-order	8
34	Self-direction	Choosing-own-goals	10
35		Creativity	13
36		Curious	10
37		Freedom	13
38		Independent	4

(continued)

Table 18.1 (continued)

#	Category	Subcategory	Number of words
39	Stimulation	A-varied-life	21
40		An-exciting-life	13
41		Daring	19
42	Tradition	Accepting-portion-in-life	6
43		Detachment	16
44		Devout	23
45		Humble	9
46		Moderate	14
47	Universalism	A-world-at-peace	18
48		A-world-of-beauty	3
49		Broad-minded	6
50		Inner-harmony	13
51		Protecting-the-environment	6
52		Social-justice	2
53		Unity-with-nature	4
54		Wisdom	32

Table 18.2 Categories of value words [20]

#	Category	Number of words
1	Power	56
2	Achievement	58
3	Hedonism	22
4	Stimulation	53
5	Self-direction	50
6	Universalism	84
7	Benevolence	119
8	Tradition	68
9	Conformity	51
10	Security	102

Table 18.3 List of actors detected with sensitivity <1.5 in all categories

#	Threshold	Actors
1	1.0	45
2	1.5	12
3	2.0	5

Table 18.4 List of actors detected with sensitivity >1.5 in all categories

#	First name	Last name	Power (%)	Achievement (%)	Hedonism (%)	Stimulation (%)	Self-direction (%)	Universalism (%)	Benevolence (%)	Tradition (%)	Conformity (%)	Security (%)
1	Susan	Pereira	0.48	0.34	0.00	0.14	0.14	0.14	0.89	0.07	0.00	1.91
2	Joe	Quenet	0.20	1.49	0.00	0.00	0.14	0.27	0.61	0.14	0.00	2.30
3	Joe	Stepenovitch	0.20	1.74	0.07	0.00	0.07	0.00	1.71	0.00	0.00	0.59
4	Harry	Arora	0.67	1.09	0.04	0.11	2.55	0.19	0.82	0.11	0.15	0.75
5	Brad	McKay	0.00	2.49	0.00	0.29	0.15	0.00	1.76	0.00	0.00	1.76
6	Tom	Donohoe	0.00	0.38	0.00	0.19	0.19	0.19	0.38	0.19	0.00	2.44
7	Mark	McConnell	0.26	0.67	0.02	0.02	0.43	0.05	0.29	0.12	0.05	1.73
8	Stacey	White	2.79	0.24	0.00	0.03	0.12	0.02	0.29	0.02	0.04	0.29
9	Phillip	Love	0.23	0.33	0.06	0.03	0.22	0.09	2.10	0.03	0.04	0.40
10	Mary	Fischer	0.30	0.43	0.04	0.00	0.47	0.22	0.82	0.00	0.04	1.72
11	Kenneth	Lay	0.81	0.40	0.51	0.00	0.00	0.40	1.82	0.10	0.10	1.62
12	Mike	Swerzbin	1.63	0.20	0.00	0.07	0.16	0.00	0.16	0.00	0.00	0.23

for each vocabulary category. The thresholds represent the percentage of the vocabulary words in each category found in the messages. As it is expected the thresholds depend on the vocabulary and are small numbers in general. Based on the specific thresholds, the Social Profile produces the list of potential insider threats. It is possible to define a common threshold for all categories or individual thresholds for each one. The profiler can then (Table 18.3) present the number of actors identified using different threshold values for any category.

Setting the threshold 1.5 for all the categories in the case study, the system identifies 12 actors under investigation. Table 18.4 presents all the scores calculated by the profiler for the specific actors.

The analyst can further check the social and behavioral profile of each identified actor by exploring the different visualizations that the tool provides. The in-depth analysis can start by looking at the overview of the actor's profile and the top statistics (see Fig. 18.3). Furthermore, the behavioral scores (see Fig. 18.4) and the word cloud (see Fig. 18.8) can provide a better indication whether the specific actor is correctly classified as a suspect. If, after the investigation, the analyst confirms that the actor is an insider threat then he/she can further explore the social profile of the identified actor and examine if there are more actors involved in the threat either as accomplices or as targets.

The analysis in the case study was done with a set of words related to ethics, however, as mentioned in the description of the profiler, the tool can operate with any vocabulary that is organized in categories and subcategories and thus it can be customized to any situation in which it could be useful. Thus, the analyst or the information security officer should select the words depending on the vocabulary of the organization e.g., the valuable assets, copyrighted materials, sensitive line of business, etc. Then by connecting the profiler with the email system and the social media of the organization, the profiler can collect the appropriate dataset and perform the necessary analysis.

Another issue, that needs to be taken seriously into consideration regarding the use of the social profiler is the security of the tool. As the tool stores in its database the communication messages of the actors and provides a meta-analysis of this information, it is understandable that it may contain private and personal information of the actors and thus access to the infrastructure where the tool is installed and to the tool itself should be restrained and always monitored. Otherwise, the tool can become a backdoor since it may reveal sensitive information to malicious users if they manage to get access to it.

Conclusions

In the current chapter, a novel social profiler tool was presented which assists the information security officer of an organization to analyze potential insider threats employing various social and behavioral analysis techniques. The tool uses a vocabulary of categories and subcategories to classify the communication of the actors involved into these categories. Based on this analysis, the results are collected and visualized using six different visualization techniques assisting the analyst to further understand the profile of the individual under investigation with the behavioral visualizations and explore his/her social neighborhood with the social profile visualizations. Following the development of the tool, its performance was tested using a known vocabulary, i.e. the set of universal values by Schwartz [20]. The communication messages of Enron employees [6] were classified against the ten different categories of the vocabulary. The identified suspicious actors were then further analyzed using the proposed visualization techniques.

The specific tool can be further enhanced by allowing the researcher to further interact with the dataset and explore more dimensions of the existing data e.g. review the adjacency matrix, compare the ego networks of more than one individual, filter the individual based on the categories they belong to and so on. Another important improvement in the tool would be the encryption of the collected data in order to avoid the accidental disclosure of information to someone with physical access to the infrastructure of the tool.

References

1. Bader BW, Berry MW, Browne M (2008) Discussion tracking in Enron email using PARAFAC. In: Survey of text mining II, pp 147–163
2. Balakrishnan B (2015) Insider threat mitigation guidance. SANS Institute InfoSec Reading Room
3. BERR (2008) Information security breaches survey 2008, 22 April 2008. http://www.eurim.org.uk/activities/ig/voi/DBERR.pdf. Accessed 17 June 2017
4. Brdiczka O, Liu J, Price B, Shen J, Patil A, Chow R, Bart E, Ducheneaut N (2012) Proactive insider threat detection through graph learning and psychological context. In: 2012 IEEE Symposium on Security and Privacy Workshops (SPW). IEEE, pp 142–149
5. Cason K (2014) Kumo—Java Word Cloud. https://github.com/kennycason/kumo. Accessed 10 May 2017
6. Cohen W (2015) Enron email dataset. Accessed 8 May 2015
7. Cole E (2015) Insider threats and the need for fast and directed response. Technical report, SANS Institute InfoSec Reading Room
8. Decherchi S, Tacconi S, Redi J, Leoncini A, Sangiacomo F, Zunino R (2009) Text clustering for digital forensics analysis. Comput Intell Secur Inf Syst 29–36
9. Eldardiry H, Bart E, Liu J, Hanley J, Price B, Brdiczka O (2013) Multi-domain information fusion for insider threat detection. In: 2013 IEEE security and privacy workshops (SPW). IEEE, pp 45–51
10. Everett M, Borgatti SP (2005) Ego network betweenness. Soc Netw 27(1):31–38
11. Fan W, Gordon MD (2014) The power of social media analytics. Commun ACM 57(6):74–81
12. Hershkop S, Stolfo SJ (2006) Behavior-based email analysis with application to spam detection. Columbia University
13. INSA (2015) Intelligence and national security alliance insider threat. https://www.insaonline.org/issues/insider-threat/. Accessed 10 June 2017
14. Kandias M, Mylonas A, Virvilis N, Theoharidou M, Gritzalis D (2010) An insider threat prediction model. In: International conference on trust, privacy and security in digital business. Springer, Berlin, pp 26–37

15. Karampelas P (2014) Visual methods and tools for social network analysis. In: Encyclopedia of social network analysis and mining. Springer New York, pp 2314–2327
16. National Institute of Standards and Technology (2014) Framework for improving critical infrastructure cybersecurity
17. Nurse JR, Buckley O, Legg PA, Goldsmith M, Creese S, Wright GR, Whitty M (2014) Understanding insider threat: a framework for characterising attacks. In: 2014 IEEE security and privacy workshops (SPW). IEEE, pp 214–228
18. Persaud A, Guan Y (2005) A framework for email investigations. In: IFIP international conference on digital forensics. Springer, US, pp 79–90
19. Poll H, Kellett A (2015) Vormetric insider threat report
20. Schwartz SH (1994) Are there universal aspects in the structure and contents of human values? J Soc Issues 50(4):19–45
21. Silowash G, Cappelli D, Moore A, Trzeciak R, Shimeall TJ, Flynn L (2012) Common sense guide to mitigating insider threats, 4th edn (No. CMU/SEI-2012-TR-012). Carnegie Mellon University Software Engineering Institute, Pittsburgh, PA
22. Spitzner L (2003) Honeypots: catching the insider threat. In: 19th annual proceedings of the computer security applications conference, 2003. IEEE, pp 170–179
23. Stolfo SJ, Hershkop S, Hu CW, Li WJ, Nimeskern O, Wang K (2006) Behavior-based modeling and its application to email analysis. ACM Trans Internet Technol (TOIT) 6(2):187–221
24. Van Alstyne M, Zhang J (2003) Emailnet: a system for automatically mining social networks from organizational email communication. Ann Arbor 1001:48109
25. Warkentin M, Willison R (2009) Behavioral and policy issues in information systems security: the insider threat. Eur J Inf Syst 18(2):101
26. Xenaros A, Karampelas P, Lekea I (2016) Profiling individuals based on email analysis and ego networks: a visualization technique. In: 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE, pp 1262–1269
27. Zhou Y (2008) Mining organizational emails for social networks with application to Enron corpus. Doctoral dissertation, Rensselaer Polytechnic Institute
28. Zhou Y, Fleischmann KR, Wallace WA (2010) Automatic text analysis of values in the Enron email dataset: clustering a social network using the value patterns of actors. In: 2010 43rd Hawaii international conference on system sciences (HICSS). IEEE, pp 1–10

Chapter 19
Cyber-Surveillance Analysis for Supercomputing Environments

A.D. Clark and J.M. Absher

Abstract High performance computers (HPCs) have contributed to rapid scientific discovery and global economic prosperity as well as defense-related applications. However, their complex nature makes them difficult to troubleshoot thus questioning their reliability. As a result, these supercomputing systems are susceptible to malicious behavior or cyber attacks. Similar investigations have been made in the context of malicious objects in computer networks; however, limited attention has been given in the context of large-scale parallel systems. In this chapter, we present a sophisticated process that characterizes observed failures in supercomputing infrastructures due to variations of consistent intentional attacks. First, we present a data network extrapolation (DNE) process that automatically does failure accounting and error checking while considering a HPC tree-like reliability infrastructure. Next, dynamic and static characterization of failures are performed. By introducing a normalization metric, we observe that the complete spectrum of failure observations is deterministic in nature that depends on the total number of failed jobs, the time between processed jobs, and the total number of processed jobs per node. Our simulations using the Structural Simulation Toolkit (SST) show that our approach is highly effective for dynamically and statically representing observed failures. Furthermore, our results can be applied for improving job-based scheduling in supercomputing environments.

A.D. Clark (✉)
Johns Hopkins University, 3400 N. Charles Street Baltimore,
Baltimore, MD 21218, USA
e-mail: adclark@mail.com

J.M. Absher
University of Maryland College Park, 8050 Greenmead
Drive College Park, Baltimore, MD 20740, USA

P. Karampelas and T. Bourlai (eds.), *Surveillance in Action*,
Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-319-68533-5_19

Introduction

High performance computers (otherwise known as HPCs) use parallel processing in order to run advanced applications efficiently and effectively that function at least a teraflop floating point operations per second (FLOPS). HPCs have been involved with many data-intensive computations in areas such as meteorology, fluid mechanics, and big data analytics. For defense applications, these systems are currently being used for command, control, communications, computers, intelligence, surveillance, and reconnaissance [1]. Data parallelism has enabled these systems to improve by several orders of magnitude [2, 3]. As a result, HPCs have accelerated technological advancements while improving overall economic prosperity. However, as the development and deployment of these supercomputing systems is continually increasing [4, 5], their reliability is in question. Furthermore, researchers have noted that their complex structure makes HPCs difficult to troubleshoot and diagnose because a single failure can affect multiple components [6]. Factors, such as these, make these systems vulnerable to cyber-attacks. Thus, advanced processes are needed to ensure systems resiliency against willful behavior.

Resilience is referred to as the ability of a system to continue operating despite the presence of faults associated with it. Thus, our goal, as stated by DeBardeleben et al. [7], is to achieve the following:

> The goal of high-end computing (HEC) resilience is to enable effective and resource-efficient use of computing systems at extreme scale in the presence of system degradations and failures.

With this in mind, recent developments in system fault detection and analysis have been made in areas such as characterizing system and application failures [6, 8] and fault-tolerant computing [9, 10]. However, limited advancements have been made to address the effects of malicious behavior resulting from cyber attacks. In this case, careful decisions need to be made with regards to protecting HPC assets while ensuring that these systems continue to be effective. If HPCs continue functioning in the midst of cyber attacks, the cyber enemy (CE) could learn about the system and submit subsequent attacks causing serious damage. Meanwhile, deciding to shut down an HPC cluster is not trivial for this task requires extensive preparation and planning [11]. Furthermore, cyber behavior can be unrecognizable as well as difficult to troubleshoot. This malicious behavior depends on the actions of the cyber-enemy (CE), which can come in various scenarios. For example, a malcontented user can access various regions of the HPC infrastructure depending on their level of access. From this, the user can introduce malicious programs that produce either job, process, or scheduling failures. These attacks can initially be camouflaged as legitimate traffic, which can later cause major system sabotage.

In this chapter, we address these concerns by proposing an overarching framework that characterizes malicious behavior in HPC systems for the case of constant attacks made anywhere within the HPC infrastructure. Our process assumes that the HPC manager will have access to the system job, fault, and error log files. Therefore, our contributions are three-fold. First, data network extrapolation (DNE) is performed that properly tallies the job-based metrics per node. Next, prediction, estimation, and analysis (PEA) process is performed that dynamically and statically characterizes system observations resulting from constant malicious attacks in terms of constant fault injection rates. Furthermore, examining the normalized fault rate and note that the spectrum of observations are sigmoidal and nonlinear regression analysis is performed to describe the behavior.

The rest of this chapter is organized in the following manner. Section "Related Work" provides a literature survey of the related work in resilience in distributed systems where we we also present our motivation and approach for handling malicious behavior in the context of HPCs. In section "HPC Cyber Surveillance Framework", we present our proposed cyber-surveillance framework. The testing process and results are presented in section "Experimentation". We conclude and discuss avenues for future exploration in section "Conclusion".

Related Work

Investigations in the area of resilience in distributed systems initially began with Von Neumann [12] who theorized physical redundancy usage for constructing highly reliable systems. His conclusion was that a redundancy of 20,000 was needed in order to achieve reliable system performance with a mean time to failure (MTTF) of 100 years. Gray later revolutionized this concept by revisiting Von Neumann's work and noting that these types of systems have migrated to modular construction [13]. As a result, he reduced the redundancy factor from 20,000 to 2 and he noted that the causes of failures can be categorized as follows: 1. software-related (50%), 2. hardware-related (30%), 3. environment-related (5%), and 4. operator-related (10–15%). Realizing that the reliability of hardware has improved, Gray later revisited his work [14] and determined that faults are predominately software-related (60%). The works of [13, 14] inspired explorations to understand failures in different systems such as telephone networks [15], networks of workstations [16], enterprise-class server environments [17], and network and internet systems [18]. In each of these cases, no generic conclusion can be made concerning the causes of failures for they are system and data dependent. For example, the work of [16] determined that failures in workstation networks are software-related. However, this conflicts with the results in [19] who conducted a similar study and demonstrated that the causes of failures are operator-related. Hence, the conclusions in [19] highlighted that planned maintenance as well as installation and configuration of software caused the largest number of outages whereas system software and planned maintenance caused the largest amount of total downtime. Explorations on this topic have also been extended to understand failures in large scale parallel systems. Motivated by the resilience analysis of large AIX clusters [20], Liang et al. [21] examined failures in IBMs BlueGene/L (BG/L) architecture where failure logs were analyzed over a period of 84 days. The authors concluded that the majority of failures in this architecture are due to network, memory, and midplane switches. Later, Schroeder and

Gibson [4] conducted an extensive study by examining failure behavior in 22 HPC systems where the conclusion was drawn that the root causes of failures are due to hardware, software, network, environment, and human operators.

Research in resilience of distributed systems has evolved from solely understanding the primary causes of failures to further examining *how* and *why* failures happen. Using a semi-Markovian failure model, Tang et al. examined the effects of error logs on a VAX cluster system and demonstrated a correlation of failure distributions amongst different machines [22]. Xu et al. [19] studied error logs from a heterogeneous system of approximately 500 personal computer (PC) servers. Their results showed that PC failures occur in bursts and rebooting systems do not completely absolve the problem. Liang et al. presented a preprocessing framework that involved a combination of three-step and temporal filtering to categorize failure events while incorporating similar errors from various locations [21]. From this strategy, an inverse relationship between the number of failure records per time unit (i.e. rate process) and the time between failures (i.e. increment process) was derived. Schroeder and Gibson's recent work illustrated that, for petascale systems, there is an inverse relationship between the expected growth in failure rates and the mean time to interrupt (MTTI) [23].

Understanding the effects of malicious behavior in distributed systems has mainly been done from the standpoint of malicious objects in computer networks. Scientists have incorporated the analogies between disease spread and computer virus propagation in order to describe virus attacks [24]. In fact, recent explorations have been made to determine the dominant effects of both viruses and worms as well as their impact when antivirus measures are considered [25]. Endemic models such as susceptible-infectious-susceptible (SIS) [26] and susceptible-exposed-infectious-recovered-susceptible (SEIRS) [27] were used to model the behavior of malicious attacks in a computer network. Additionally, worm transmission models were also studied. Mishra and Pandey represented the propagation of worms in a computer network via a susceptible-exposed-infectious-susceptible with vaccination (SEIS-V) framework [28]. Their work involved deriving an explicit equation for quantifying the spread of infected nodes while incorporating the effects of an efficient virus software. Furthermore, modeling worm propagation has also been explored in other domains and topologies where examples include peer-to-peer (P2P) networks [29], Facebook and email networks [30], and mobile network topologies [31].

Although there have been many investigations concerning malicious activity in computer networks, limited attention has been given in terms of nefarious behavior in large scale parallel systems. Faissol and Gallagher [32], motivated by the works of [33, 34], employed game theoretic approaches to examine human computer interaction (HCI) within HPC environments. By modeling the situation as a discrete zero sum stochastic game, they numerically analyzed this impact via determining the trade-offs between the *price of anarchy (POA)* and *price of malice (POM)*. Here, the *price of anarchy (POA)* is the impact of selfish users and the *price of malice (POM)* is the impact of malicious users [33]. With this approach the authors were able to characterize the systems ability of detection, the relative cost of attacks and repairs, the game length, and background system characteristics. With the goal of improving the prediction and representation of observed failures resulting from intentional attacks, Pritchett-Sheats used the conditional maximum likelihood estimator (CMLE) while considering treelike models of HPC infrastructures [35]. For small configurations, the author concludes that the CMLE method does well when it comes to predicting and characterizing observed failures. Motivated by the work of Faissol and Gallagher [32], Clark [36] revisited the work of [32] and analytically modeled the HCI interactions as a composite Markov process. The results of his model were consistent with the work of [32] for the case of constant attacks. For constant switching attacks, the work in [36] analytically showed that the malicious impact in the overall system directly depends on the cyber attack rates between failure risk levels and the overall system failure rate. Motivated by the work of [35], Clark et al. [37] adopted the HPC reliability structure in [35] and produced an approach to dynamically forecast failures for each node. Their results, showed significant improvement in terms of predicted and observed failures per day with an earlier prediction by at least two orders of magnitude in terms of total number of jobs processed per node.

Our Motivation and Approach

Our motivation is from the works of [35, 37] that illustrate the need for robust processes that dynamically and statically forecast, define, and evaluate malicious behavior within HPCs. Pritchett-Sheats notes that although employing the CMLE approach shows promise for small-scale architectures, this approach also showed diminishing returns. As HPC configurations and job sizes increased, the CMLE estimator actually depends on the law of large numbers where at least $1,000$ jobs are needed to perform better prediction. Furthermore, the approach employed in [35] considers failures across a subregion within the HPC environment. Therefore, the limitations posed in [35] can be problematic because improperly performing failure accounting impacts prediction performance. And although the work of [37] showed much promise in terms of adaptive forecasting, additional processes are needed to evaluate the complete spectrum of intentional behavior.

Our work extends that of [35, 37] where we propose a cyber resilience schema, shown in Fig. 19.1, that statically and dynamically characterizes malicious behavior per node where we use the same tree reliability network. Here, we assume that the cyber-enemy can access anywhere within the HPC infrastructure and perform constant injections. Our analysis centers on the 15-node reliability network, which is motivated by the treelike structure proposed in [6]. First, we propose a data network extrapolation (DNE) process calculates the complete statistics in terms of total number of successful, failed, and total jobs processed while preserving the nodule dependencies in the HPC reliability infrastructure. Next, our prediction, estimation, and analysis (PEA) module performs the static and dynamic failure rate characterization. Our dynamic characterization and prediction process involves *spline windowing* where only a modicum of observations are used to predict the next failure. Our static characterization process is based on our observation that the spectrum of

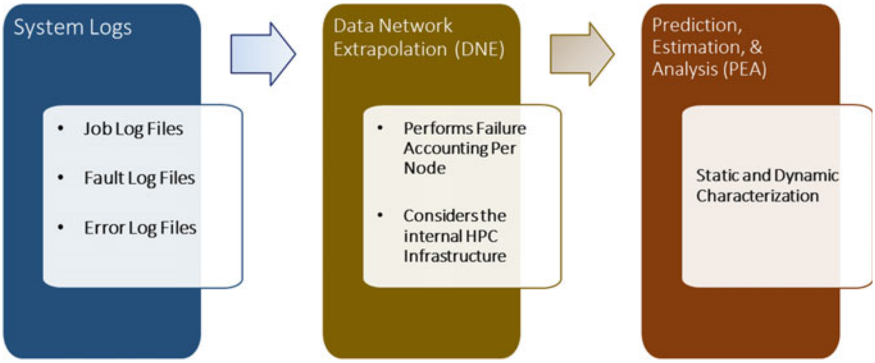

System Logs
• Job Log Files
• Fault Log Files
• Error Log Files
Data Network Extrapolation (DNE)
• Performs Failure Accounting Per Node
• Considers the internal HPC Infrastructure
Prediction, Estimation, & Analysis (PEA)
Static and Dynamic Characterization

Fig. 19.1 Overview of our proposed cyber surveillance framework

failures are both deterministic and sinuous in nature that depend on the maximum and minimum failure rates as well as the changes in failure rate per number of jobs processed where nonlinear regression is performed to characterize this behavior.

HPC Cyber Surveillance Framework

Our cyber surveillance methodology is based on the following assumptions. First, we assume that within the HPC infrastructure the collections of systems and processes are dependent on one another in the form of a directed acyclic graph where Fig. 19.2 serves as an example. We also assume that the HPC will produce the following files:

- *Job Log Files*—hese files provide the job status as to whether a job succeeded, failed or postponed within the reliability infrastructure. This information is only provided at the lowest end of the reliability network structure.
- *Fault Log Files*—These files provide information as to where and when faults occur within the reliability network.
- *Error Log Files*—These files, which are extensions of fault log files, provide the fault dependencies within the system.

It is assumed that failures can occur at any location within the HPC reliability infrastructure to consider possibilities where malicious algorithms can be injected. Furthermore, we do not consider any type of propagation between nodes and failure and job processing checkpoints can be set at any node at the discretion of the HPC manager.

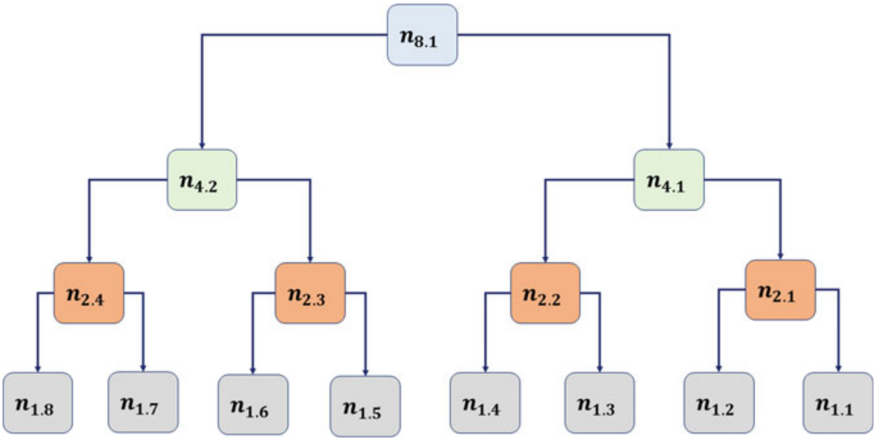


Fig. 19.2 Example representation of a reliability infrastructure within a HPC environment where each node is representative of either single or a collection of processes

Data Network Extrapolation (DNE) Process

Taking these assumptions into account, we first determine the complete metrics at each component to include the total number of successful and failed jobs as well as the total number of jobs processed. This provides an accurate calculation of the failure rate. Examining the reliability network in Fig. 19.2, if an attack produced a failure at node $n_{2.1}$ then failures would be produced at $n_{2.1}$ as well as its dependent compute nodes $n_{1.1}$ and $n_{1.2}$. This is true for any hierarchal node within the network. However, attacks and failures are directly associated for any compute node because these nodes occur at the lowest level in the reliability hierarchy. Therefore, computing the total metrics involves knowing the original location of failures within the HPC reliability network. Otherwise, as more jobs are scheduled and processed, the metrics across each node become progressively erroneous affecting the characterization of failure rates.

Our data network extrapolation (DNE) process overcomes these limitations via accurately accounting for the error metrics at each node while taking into account the origination of failures. The DNE process serves as a preprocessing methodology that incorporates the system log files and calculates the total number of failed and successful jobs at each node as well as the total number of jobs processed (Fig. 19.3). The data extraction is performed via traversing through the system information while considering the dependency relationships. For example, if a failure is recorded at nodes $n_{1.1}$, $n_{1.2}$, and $n_{2.1}$, then it can be inferred that the failure originated at node $n_{2.1}$. We continue this process until all of the jobs are completed. As a result, we are able to determine the complete metrics per node. Hence, we calculate the observed nodule fault rate λ_o given by

$$\lambda_o = f/\Delta t, \tag{19.1}$$

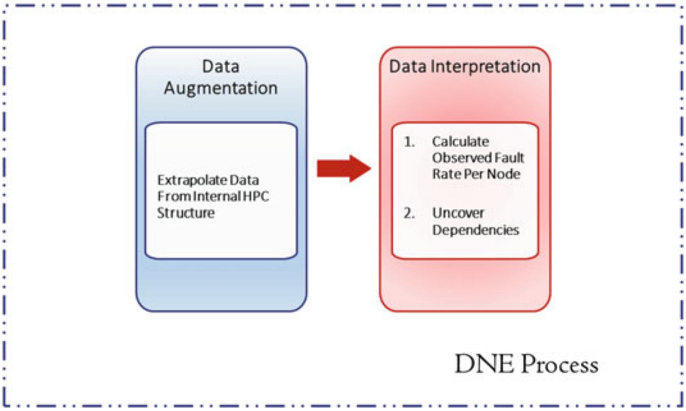

Data Augmentation
Extrapolate Data From Internal HPC Structure
Data Interpretation
1. Calculate Observed Fault Rate Per Node
2. Uncover Dependencies
DNE Process

Fig. 19.3 Overview of our proposed data network extrapolation (DNE) process

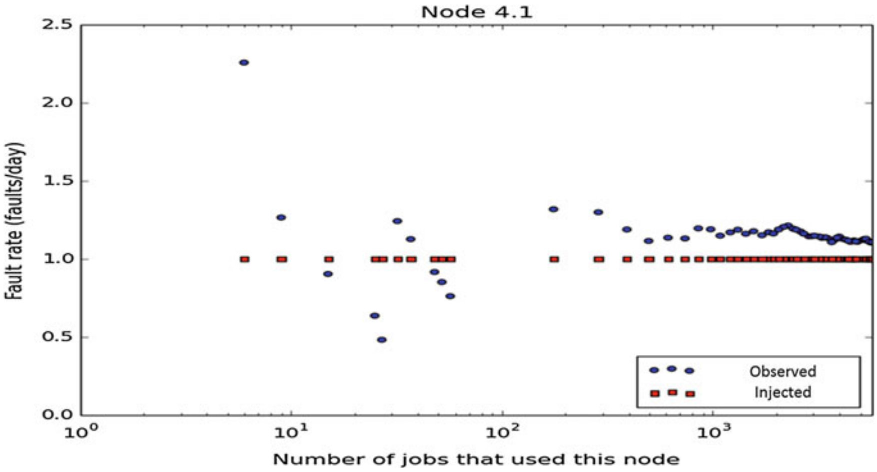


Fig. 19.4 Example of a simulated response of observations (in terms of failure rate) when an intentional (or injected) fault rate of 1.0 is injected into node $n_{4.1}$ given the HPC reliability structure shown in Fig. 19.2

where f is the total number of faults (per node) and Δt is the time differential between job processes. Now, we are able to graph the true observed response resulting from an intentional fault injection where an example is shown in Fig. 19.4.

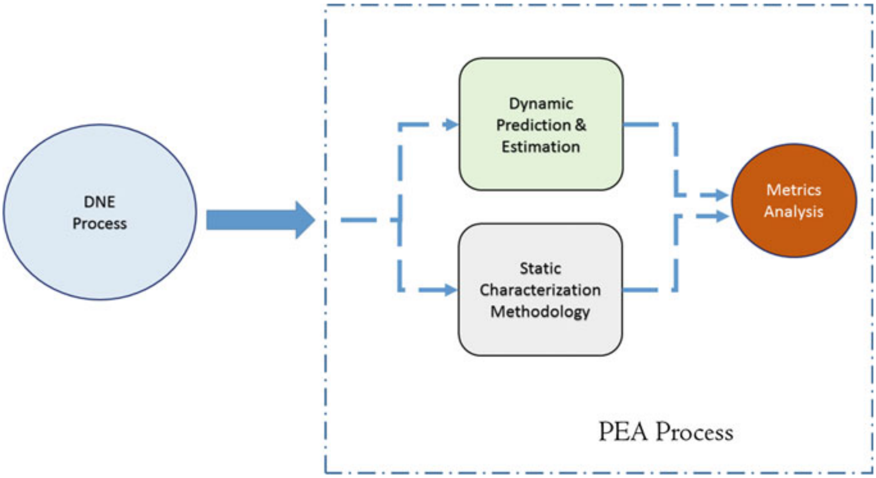

DNE Process
Dynamic Prediction & Estimation
Static Characterization Methodology
Metrics Analysis
PEA Process

Fig. 19.5 Overview of our proposed prediction, estimation, and analysis (PEA) process

Prediction, Estimation and Analysis (PEA) Methodology

One might intuitively assert that the observations and the injected failures are identical. However, as shown in Fig. 19.4, this is not the case for the observations depend on the following factors: 1. total number of jobs processed, 2. total number of successes, and consequently 3. total number of failures. Therefore, it is also important to depict our observations. Hence, we also incorporate a prediction, estimation, and analysis (PEA) process, illustrated in Fig. 19.5, that characterizes the resulting observations statically and dynamically. Static interpretation is needed in order to learn about the spectrum of failure behavior. At the same time, adaptive forecasting is needed to anticipate malicious behavior while engineering preventative measures.

Dynamic Characterization

Our dynamic characterization process involves employing windowing via smoothing splines (otherwise known as *smooth spline windowing*) via the following expression

$$\sum_{j=1}^{N} \left(Y_j - \hat{\mu}\left(x_j\right)\right)^2 + \alpha \int_{x_1}^{x_N} \hat{\mu}''(\rho) d\rho, \tag{19.2}$$

where for each observation, modeled as $Y_j = \mu\left(x_j\right)$, $\hat{\mu}$ is spline estimate,[1] and $\alpha \geq 0$ is the smoothing parameter. The second term, in Eq. (19.2), is known as the regularization term that is capable of shrinking the estimations $\hat{\mu}$ where α is the tuning parameter. When $\alpha \to 0$ (i.e. when no smoothing is imposed), Eq. (19.2) becomes a typical interpolating spline. However, for the case of infinite smoothing (i.e. as $\alpha \to \infty$), the smoothing spline becomes a linear least squares estimate. Smoothing splines are valuable for they provide the following advantages: 1. improved flexibility, 2. better knot selection, and 3. better over-fitting control.

Our windowed process involves a length of four consecutive observations where the initial set is interpolated, using (19.2), to provide a historical baseline. Next, we predict while keeping the interpolation region small within six observations in order to track the forecast of the next observation. Our analysis is done on the fly by making a comparison between the subsequent predicted and observed failure rates in terms of the residual squared error (RSE) to dynamically track performance. This marching process continues until either all of the observations have been exhausted or the process is terminated by the end user.

Static Characterization

As shown in Fig. 19.6, we observe that the change in successful and failed jobs, and consequently the observed failure rate, is correlated between the fault injection rate and the total number of processed jobs. We also note that Eq. (19.1) does not account for periods of inactivity. Thus, we introduce the *normalized* observed fault rate $\bar{\lambda}_o$ given by

$$\bar{\lambda}_o = \frac{f/\Delta t}{N_J} = \frac{\lambda_o}{N_J}, \tag{19.3}$$

where the observed fault rate λ_o is normalized by the total number of jobs processed per node N_J. This spectrum of observations is sigmoidal or “S”-shaped, which can be represented via the generalized logistic function given as

$$\bar{\lambda}_o(x) = \beta_1 + \frac{\beta_2 - \beta_1}{\left[1 + \beta_3 e^{\beta_4\left(\log_{10}(x) - \beta_M\right)}\right]}, \tag{19.4}$$

where x represents the system checkpoint, in terms of number of jobs processed, $\bar{\lambda}_o$ is the associated normalized failure rate, and the constants β_1 through β_M are defined in Table 19.1. An equivalent four parameter model to Eq. (19.4) can be defined by allowing $\beta_3 = 1$.

This process computes the parameters β_1 to β_M of Eq. (19.4) given the spectrum of system checkpoints. To do this, we first perform the initial estimates $\hat{\beta}_1$ to $\hat{\beta}_M$.

[1] The spline estimates are third degree polynomials.

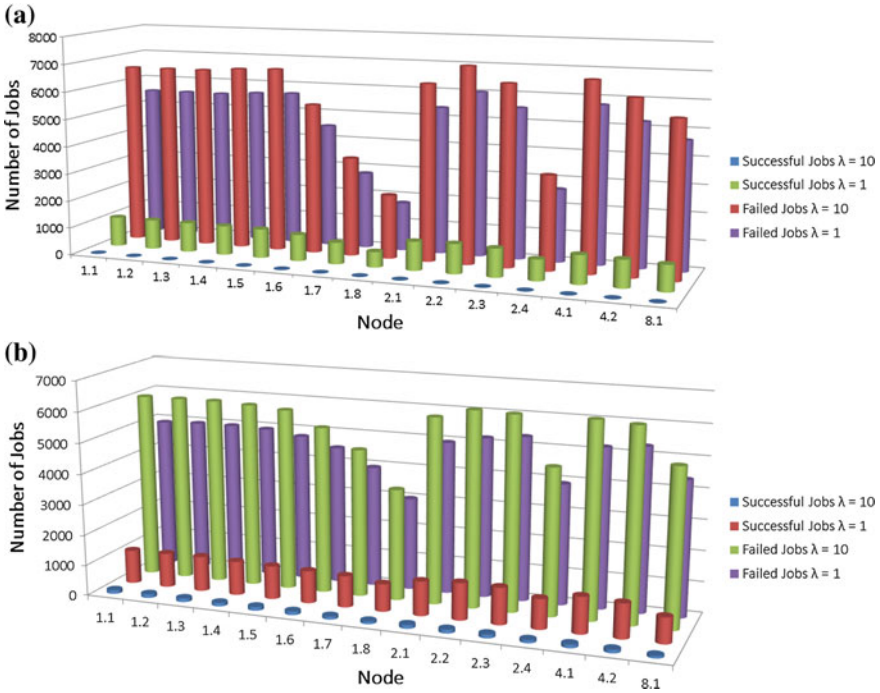


Fig. 19.6 Simulation of the impact of injected attacks λ within each node of the HPC reliability infrastructure while processing 10,000 jobs. Our comparisons are made for the cases where $\lambda = 1$ and $\lambda = 10$ faults per day for the following cases: **a** deterministic job sizes and durations and **b** when the job sizes and durations are completely randomized

Table 19.1 Definition of constants for Eq. (19.4)

Constant	Definition
β_1	Lower asymptote
β_2	Upper asymptote
β_3, β_M	Determines midpoint behavior
β_4	Growth rate

Observing the monotonicity of $\bar{\lambda}_o(x)$, the initial estimates $\hat{\beta}_1$ and $\hat{\beta}_1$ are determined via employing the quantile function $Q_{\bar{\lambda}_o}$ defined as

$$Q_{\bar{\lambda}_o}(p) = \inf\left\{x \in R\left(\bar{\lambda}_o\right) : p \leq \bar{\lambda}_o(x)\right\}, \tag{19.5}$$

where $R\left(\bar{\lambda}_o\right)$ is the range of $\bar{\lambda}_o$ with probability $p \in (0, 1)$. From Eq. (19.5), the initial estimates of the lower and upper asymptotes are found to be $\hat{\beta}_1 = Q_{\bar{\lambda}_o}(1 - p_a)$ and $\hat{\beta}_2 = Q_{\bar{\lambda}_o}(p_a)$ where $p_a \in (0, 1)$. $\hat{\beta}_M$ is estimated as the median of the domain

$\log_{10}(x)$. By performing the following normalization

$$z = \frac{\bar{\lambda}_o(x) - \hat{\beta}_1}{\hat{\beta}_2 - \hat{\beta}_1}, \tag{19.6}$$

$\hat{\beta}_3$ is estimated via the relationship

$$\hat{\beta}_3 = 1/\zeta_M - 1 \tag{19.7}$$

where ζ_M is the median estimate of the values resulting from Eq. (19.6). Next, the initial growth rate estimate $\hat{\beta}_4$ is found by performing a linear fitting to the equation

$$\hat{\beta}_4 \left(y - \hat{\beta}_M\right) = \ln\left(\frac{1-z}{z}\right) - \ln\left(\hat{\beta}_3\right) \tag{19.8}$$

where $y = \log_{10}(x)$ and z is defined by Eq. (19.6). This fitting determines the slope, which represents the initial valuation of the growth rate.

Now we follow the approach in [38] in finalizing the parameters β_1 to β_M from the data generated from the DNE process. Letting $\mathcal{B} = (\beta_1, \beta_2, \dots, \beta_M)$ be the vector of parameters to be estimated, the goal is to minimize the sum of squares residual R_{SS} given by

$$R_{SS} = \sum_{i=1}^{N} \left[\omega_i - \bar{\lambda}_o(x_i, \mathcal{B})\right]^2 \tag{19.9}$$

where x_i and ω_i represent each system checkpoint and associated observation, respectively. Hence, it suffices to solve the following equation

$$\sum_{i=1}^{N} \left\{\omega_i - \bar{\lambda}_o(x_i, \mathcal{B})\right\} \left[\frac{\partial \bar{\lambda}_o(x_i, \mathcal{B})}{\partial \beta_i}\right] = 0 \tag{19.10}$$

where β_i is each individual parameter, β_1 to β_M, to be estimated.

Metrics Analysis

From the static and dynamic characterization, metrics analysis is performed to determine the following:

- *Dynamic Characterization Effectiveness*—To evaluate how well our dynamic characterization aligns with the true observation as well as to determine the level of adaptiveness as each observation is realized.

- *Static Characterization Effectiveness*—To evaluate how closely our static characterization depicts the spectrum of observations within each node of the reliability network.

The *dynamic characterization effectiveness* is quantified by comparing the root mean squared error (RMSE) to the mean absolute error (MAE) of each node to determine the level of variation within the errors. Their difference in magnitude, ε = RMSE − MAE, is also used to understand the error difference. The greater the difference between the MAE and the RMSE the greater the variance between the predicted and true observations. The level of *static characterization effectiveness* is centered on examining the goodness of fit (GoF) to determine how well Eq. (19.4) aligns with the range of observations.

Experimentation

Our performance evaluations were done using the Structural Simulation Toolkit (SST)—an open-source, modular, parallel, multi-objective, and multi-scale simulation framework for HPC exploration [39]. For our purposes, SST is used to impose failures within several components of the HPC reliability network shown in Fig. 19.2. The goal is to evaluate the effectiveness of our proposed strategy via examining the metrics of the DNE and PEA processes. In evaluating the DNE process the objective is to ensure that for each node the relationship between the total number of successful, failed, and processed nodes are properly accounted. This ensures that the observed failure rate is accurately calculated. The aim for evaluating the PEA process is to determine the level of fidelity in the dynamic and static characterization modules.

Our testing strategy involves injecting various constant failure rates throughout all of the nodes of the HPC system configuration shown in Fig. 19.2 where we first performed baseline assessments between the injected and observed failure rates. Next, we considered different variations in job durations and sizes where the average job time is approximately 6 hours. We varied the job sizes incrementally in increments of 10 starting at 10 jobs to 100 jobs followed by increments of 200 from 100 jobs to 9, 900 jobs. In an actual HPC environment the scheduling of both job sizes and durations are random; however, we needed to ascertain baseline performance. Therefore, as Table 19.2 suggests, we gradually introduce random behavior with both job durations and sizes while performing deterministic variations in both categories.

DNE Process Results

Table 19.3 illustrates the fidelity of our proposed DNE process, which shows detailed accounting per node. The total number of successful and failed jobs are directly

Table 19.2 Testing methodology

Job duration	Job size
Decreasing	Decreasing
Decreasing	Increasing
Decreasing	Random
Increasing	Decreasing
Increasing	Increasing
Increasing	Random
Random	Decreasing
Random	Increasing
Random	Random

Table 19.3 DNE accounting per node

Node	Total jobs	Total successes	Total failures
8.1	5267	5	5262
4.2	6613	52	6561
4.1	6083	32	6051
2.4	4788	206	4582
2.3	6690	75	6615
2.2	6175	48	6127
2.1	5857	41	5816
1.8	4387	506	3881
1.7	4845	232	4613
1.6	5512	97	5415
1.5	6043	64	5979
1.4	6015	57	5958
1.3	5912	69	5843
1.2	5911	66	5845
1.1	5911	65	5846

correlated with the total number of jobs processed per node, which are consistent regardless of the total number of scheduled and processed jobs. These metrics further show the robustness of our accounting methodology, which leads to the proper failure rate estimation. Thus, serving as a preprocessing component to our PEA process.

Table 19.4 Performance metrics (failures/day)

Node	MAE	RMSE	ε = RMSE − MAE
8.1	0.32035	0.44451	0.12416
4.2	0.22598	0.27485	0.04887
4.1	0.26228	0.36413	0.10186
2.4	0.24152	0.32031	0.07879
2.3	0.28812	0.34511	0.05700
2.2	0.17580	0.22236	0.04656
2.1	0.30836	0.52660	0.21824
1.8	0.28197	0.40646	0.12449
1.7	0.29901	0.55640	0.25739
1.6	0.26652	0.34386	0.07734
1.5	0.34415	0.74257	0.40142
1.4	0.27819	0.35797	0.07978
1.3	0.27449	0.35644	0.08195
1.2	0.27851	0.36765	0.08014
1.1	0.20954	0.30794	0.09840

PEA Process Results

Table 19.4 presents the advantage of employing spline windowing in the dynamic characterization where the mean difference between forecasted and realized observations in the range of $(0.17, 0.35)$ failures per day. Although the root mean squared error (RMSE) is slightly higher, the difference $\varepsilon = \text{RMSE} - \text{MAE}$ is consistently small, with $\varepsilon < 1$, providing further validation of our proposed process. The results of the RMSE values at nodes 1.5, 1.7, and 2.1 might bring to question the validity of our process; however, it must be noted that this is in terms of failures per day. In other words, our adaptable process is able to anticipate failure rate observations within one failure per day when approximately $10,000$ jobs have been processed. It is important to note that our simulations show that this technique achieves an interpolation penalty of $O(10^{-23})$ at 50 degrees of freedom with a convergence of approximately four observations [37]. Therefore, in addition to faster convergence our proposed dynamic process also provides more accurate forecasting.

Table 19.5 provides a comparison in terms of goodness of fit (GoF) when four ($\beta_3 = 1$) and five parameter representations of the generalized logistic model are used when $9,900$ jobs are processed for each node. By examining the GoF with the weighted residual for each node, we empirically determined the cut off for the GoF is 0.70 ± 0.05. Hence, our results show that in both cases our static characterization via Eq. (19.4) does well in modeling the assortment of observations for approximately 70% of the nodes.

Table 19.5 Static characterization results (Four Parameter Fitting)

Node	Goodness of fit (GoF) (four parameter)	Goodness of fit (GoF) (five parameter)
8.1	0.986	0.998
4.2	0.877	0.971
4.1	0.504	0.962
2.4	0.708	0.459
2.3	0.962	0.949
2.2	0.394	0.356
2.1	0.132	0.129
1.8	0.528	0.573
1.7	0.652	0.687
1.6	0.950	0.821
1.5	0.948	0.941
1.4	0.653	0.695
1.3	0.220	0.518
1.2	0.878	0.977
1.1	0.891	0.889

Conclusion

This chapter presents a cyber surveillance framework for determining and characterizing malicious behavior in supercomputing infrastructures consisting of data network extrapolation (DNE) and prediction, estimation, and analysis (PEA) modules. The DNE process serves as a preprocessing schema where failure rate determinations are made per node of the reliability network, shown in Fig. 19.2. This is followed by the PEA process where dynamic and static characterization is performed. Dynamic characterization is important for tracking failure behavior to continually assess malicious behavior. Static characterization is also imperative for assessing the overall assortment of resulting failure observations per node.

Our empirical results illustrate promise in our overall framework. The results of our DNE process demonstrate the robustness of our failure accounting. The dynamic characterization results of our PEA process show that, for each node, our predictions are within one failure rate per day. Additionally, when compared to the results in [35], our predictions only require three observations resulting in an improvement by at least an order of magnitude. Our static characterization results were also effective, which showed a strong representation for approximately 70% of the nodes. These results, individually and collectively, show robustness in our overall framework compared to prior approaches that rely solely on the law of large numbers for failure rate characterization.

The results of this work have several applications within the supercomputing resilience community. Our schema can be applied to characterize and diagnose intentional behavior in some HPC environments, like the *Cray XC Series* [40], that contain treelike reliability networks. Specifically, our framework can be used to troubleshoot the location of these occurrences while providing alternative strategies to mitigate malicious activity. For example, the dynamic characterization module can be applied as a detection and correction mechanism for proper rescheduling to avoid interruptions in supercomputing processing. Our static characterization process can be used to troubleshoot, diagnose, and classify cyber anomalies where comparisons can be made in terms of the spectrum of failures between nodes. From a research perspective, our work can be used to comprehensively understand the impact of different types of cyber activity within a supercomputing environment. For example, [32] employs game theoretic approaches to understand risk level behavior; however, as mentioned in [36], it is important to understand both the dynamics and the constraints of the behavior. Our work can be employed to determine certain scenarios of intentional failures where game theory can be applied. Additionally, our work can be applied to [36] to provide estimations of the probabilistic actions to access risk level behavior.

Although our approach is an advancement over prior methods, there are avenues of future work that need to be explored. First, it is important to note that the reliability network proposed in Fig. 19.2 is a simplistic structure. Therefore, investigations need to be made for treelike networks beyond the 15-node structure. Another area of exploration is extending this process to test for cases of varying actions. An example of this would be the case where malicious behavior would consist of 10 faults per day only at certain nodes while other nodes would have an injection rate of approximately 1 fault per day. For this case, the concept of variegated fault injection rates can also be explored. Furthermore, there is a need to determine various classes of targeted behavior along with their depictions. Conducting these investigations, along with surveying current and previous approaches, will help develop comprehensive perspectives of this important and intriguing phenomenon.

Acknowledgements This work is based upon work supported by the U.S. Army Research Laboratory (ARL) under award number W911NF-13-D-0006. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory. We would also like to thank Dr. Tara Michels-Clark and the anonymous reviewers for their insightful comments and suggestions.

References

1. Howard C (2011) Military & aerospace electronics
2. Nyland LS, Prins JF, Goldberg A, Mills PH (2000) IEEE Trans Software Eng 26(4):293
3. Ravichandran D, Pantel P, Hovy E (2004) KDD workshop on mining for and from the semantic web (MSW-04). Citeseer, pp 1–11
4. Schroeder B, Gibson GA (2007) J Phy Conf Ser, vol 78. IOP Publishing, pp 12–22

5. Chen Z, Dongarra J (2008) IEEE Trans Parallel Distrib Syst 19(12):1628
6. Daly JT, Pritchett-Sheats LA, Michala SE (2008) 8th IEEE international symposium on cluster computing and the grid (CCGRID). IEEE, pp 795–800
7. DeBardeleben N, Laros J, Daly J, Scott S, Engelmann C, Harrod B (2009) Whitepaper
8. Jones WM, Daly JT, DeBardeleben NA (2008) 8th IEEE international symposium on cluster computing and the grid. IEEE, pp 789–794
9. Raicu I (2009) Many-task computing: bridging the gap between high-throughput computing and high-performance computing. ProQuest
10. Lunacek M, Braden J, Hauser T (2013) IEEE international conference on cluster computing (CLUSTER). IEEE, pp 1–8
11. Quintero D, Bosworth K, Chaudhary P, da Silva RG, Ha B, Higino J, Kahle ME, Kamenoue T, Pearson J, Perez MM et al (2014) IBM power systems 775 for AIX and Linux HPC solution. IBM Redbooks
12. Neumann JV (1956) Autom Stud 34:43
13. Gray J (1986) Symposium on reliability in distributed software and database systems, pp 3–12
14. Gray J (1990) Reliability. IEEE Trans 39(4):409
15. Kuhn RD (1997) Computer 30(4):31
16. Kalyanakrishnam M, Kalbarczyk Z, Iyer R (1999) 18th IEEE symposium on reliable distributed systems. IEEE, pp 178–187
17. Lee I, Iyer RK (1995) IEEE Trans Softw Eng 21(5):455
18. Oppenheimer D, Ganapathi A, Patterson DA (2003) USENIX symposium on internet technologies and systems, vol 67. Seattle, WA
19. Xu J, Kalbarczyk Z, Iyer RK (1999) Pacific rim international symposium on dependable computing. IEEE, pp 178–185
20. Sahoo RK, Squillante MS, Sivasubramaniam A, Zhang Y (2004) International conference on dependable systems and networks. IEEE, pp 772–781
21. Liang Y, Zhang Y, Sivasubramaniam A, Sahoo RK, Gupta JMM (2005) International conference on dependable systems and networks. IEEE, pp 476–485
22. Tang D, Iyer RK, Subramani SS (1990) 20th international symposium on fault-tolerant computing. IEEE, pp 244–251
23. Schroeder B, Gibson GA (2010) IEEE Trans Dependable Secure Comput 7(4):337
24. Yuan H, Chen G (2008) Applied Math Comput 206(1):357
25. Data S, Wang H (2005) Canadian conference on electrical and computer engineering. IEEE, pp 219–223
26. Kim J, Radhakrishnan S, Jang J (2006) ETRI J 28(5):692
27. Mishra BK, Jha N (2010) Appl Math Model 34(3):710
28. Mishra BK, Pandey SK (2014) Appl Math Model 38(7):2173
29. Chen TM (2013) J Netw Comput Appl 36
30. Fan W, Yeung K (2013) The influence of technology on social network analysis and mining. Springer, pp 185–199
31. Wang P, González MC, Menezes R, Barabási A (2013) Int J Inf Secur 12(5):383
32. Faissol G, Gallagher B (2014) Lawrence livermore national laboratory (technical report)
33. Moscibroda T, Schmid S, Wattenhofer R (2006) 25th annual ACM symposium on principles of distributed computing. ACM, pp 35–44
34. Koutsoupias E, Papadimitriou C (2009) Comput Sci Rev 3(2):65
35. Pritchett-Sheats LA (2013) Los Alamos national laboratory (technical report)
36. Clark AD (2016) Handbook of research on next-generation high performance computing. IGI Global
37. Clark AD, Tellez LM, Besse S , Absher JM (2016) IEEE/ACM international conference on advances in social networks analysis and mining. IEEE/ACM
38. Fekedulegn D, Siurtain MPM, Colbert JJ (1999) Silva Fennica
39. Rodrigues AF, Hemmert KS, Barrett BW, Kersey C, Oldfield R, Weston M, Risen R, Cook J, Rosenfeld P, CooperBalls E et al (2011) ACM SIGMETRICS Perf Eval Rev 38(4):37
40. Schmidtke R, Laubender G, Steinke T (2016) Cray User Group (CUG)