Gaming Media and Social Effects

Yiyu Cai
Sui Lin Goei *Editors*

Simulations, Serious Games and Their Applications

Springer

Gaming Media and Social Effects

For further volumes:
http://www.springer.com/series/11864

Yiyu Cai · Sui Lin Goei
Editors

Simulations, Serious Games and Their Applications

Springer

Editors
Yiyu Cai
Nanyang Technological University
Singapore

Sui Lin Goei
Windesheim University of Applied Science
Bloemendaal
The Netherlands

ISSN 2197-9685 ISSN 2197-9693 (electronic)
ISBN 978-981-4560-31-3 ISBN 978-981-4560-32-0 (eBook)
DOI 10.1007/978-981-4560-32-0
Springer Singapore Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013953197

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

In early 2012, I had a chance to visit Singapore. Introduced by my colleague Prof. Sui Lin Goei, a visiting professor with Singapore's National Institute of Education at that time, I met with Prof. Yiyu Cai at the Institute for Media Innovation, Nanyang Technological University in Singapore. During the meeting, we discussed potential collaboration in the field of Serious Games and Simulation. As one of the follow-ups, the Asia–Europe Workshop on Simulation and Serious Games was held in Nanyang Technological University on 9 May 2012, as part of the 23rd Conference on Computer Animation and Social Agents. The workshop attracted good participants from both Asia and Europe. I led a team from The Netherlands, attended the workshop, and had a further discussion with Prof. Cai and his colleagues on collaboration between Singapore and The Netherlands. This book edited by Prof. Cai and Prof. Goei includes several interesting research presented at the workshop.

As Dean of the Faculty of Education, Windesheim University of Applied Sciences, I am well aware of the impact that Computer Simulations and Serious Gaming can have on education. With the introduction of ever more powerful computers, Serious Gaming has become one of the fastest growing areas in digital media. Currently, education is still only at the very start of exploring the possibilities of Serious Gaming. Therefore, next to the technological developments there is an increasing need to investigate the efficient and effective use of Serious Gaming in educational programmes. An important issue in this research will be the behaviour of those who are learning by using Serious Gaming. This book will provide an insight into these developments and into the possibilities that Computer Simulations and Serious Gaming have to offer to education.

The next big challenge for education will be to further develop and investigate the possible applications within curricula and didactical concepts. Hence, I do see many opportunities to apply Computer Simulations and Serious Gaming in remedial teaching, distance learning, specific skills programmes, honours programmes, behaviour training programmes, and so on. To head in this way successfully depends on international cooperation. In Asia various countries

possess technological know-how. In Europe there is a lot of know-how about educational innovation. Combining these disciplines will make an adequate development and use of Serious Gaming within education possible. I will make a strong effort to stimulate cooperation between Asia and Europe through means of joint projects.

Please let this book be a big inspiration to you as it was to me.

The Netherlands Harry Frantzen

Preface

Driven by applications, Serious Game is one of the fast growing fields in the Interactive and Digital Media (IDM). As an enabling technology, Simulation plays a crucial role in Serious Games. While Simulation research has a long history, Serious Game study, however, is still in its infant stage. On 9 May 2012, researchers and developers from two continents came together to attend The 2012 Asia–Europe Workshop on Simulation and Serious Games which was held in Nanyang Technological University, Singapore.

This book describes the new development in Simulations and Serious Games selected mostly from the above-mentioned workshop. Collision detection is a fundamental issue in Simulation and Serious Games. Panpan Cai et al. and Xinyu Zhang will discuss their collision detection algorithms and simulation applications. For Chinese calligraphy ink diffusion simulation, Yuzhi Ren et al. will present a physically based method using Computational Fluid Dynamics. For cloth simulation, Yuzhe Zhang et al. will describe the human body adaptation and cloth prepositioning for Kinect-based virtual try-on. For occupational therapy of mental patients, Chun Siong Lee et al. will postulate their solution by combining human performance engineering with cognitive science for effective autonomous rehabilitative training. For marine traffic conflict control, Qing Li et al. will propose their algorithm and simulation model to predict likely conflicts with multi links ahead before vessels actually encounter. For long vehicle turning, Yong Chen et al. will elaborate their modelling and simulation for trajectory planning for long vehicle tuning by solving differential equations. For Virtual Worlds and Serious Gaming in Education, Wim Trooster will share his experience with the Second Life. For health care, Voravika Wattanasoontorn et al. will investigate serious games for the application of e-Healthcare services. For game design and development, Hock Soon Seah et al. will detail a versatile platform—The GF Engine, and Rubén Jesús García et al. will report their research on updated GameTools and application in serious gaming. For engineering training, Peiling Liu et al. will show their CNC simulation system research and development. For special needs

education, Noel Chia et al. will present their serious and simulation game of virtual pink dolphins. Sandra Tan will share their experience of using simulations to improve visual spatial ability in molecular biology learning. E. A. P. B (Esther) Oprins and J. E (Hans) Korteling will discuss the effectiveness of the transfer of training with the serious game—Cashier Trainer.

Yiyu Cai
Sui Lin Goei

Contents

Collision Detection Using Axis Aligned Bounding Boxes

Panpan Cai, Chandrasekaran Indhumathi, Yiyu Cai, Jianmin Zheng, Yi Gong, Teng Sam Lim and Peng Wong

Abstract Collision detection plays a critical role in real-time applications such as game, simulation, and virtual reality. Collision avoidance is important in robotics path planning. Industrial safety, especially in construction and building, has a close linkage with the concept of contact avoidance. This chapter is interested in the investigation of collision detection problem using hardware graphics acceleration. Axis aligned bounding boxes (AABB) technique will be applied also for fast collision detection.

Keywords Axis aligned bounding boxes · Collision detection · Simulation

1 Introduction

Collision detection plays a critical role in real-time applications such as game, simulation, and virtual reality. Collision avoidance is important in robotics path planning. Industrial safety, especially in construction and building, is closely related to contact avoidance. High-accuracy and real-time collision detection for physically based simulations is often computationally expensive (Fauer et al. 2008). Apart from making the simulation more intelligent and realistic, the collision detection engine is also the basis of many applications such as collision response and haptic feedback.

Usually, collision detection algorithms are divided into two phases: the *broad phase* and the *narrow phase* (Govindaraju et al. 2003). The broad phase aims at

P. Cai · C. Indhumathi · Y. Cai (✉) · J. Zheng
Nanyang Technological University, Singapore, Singapore
e-mail: myycai@ntu.edu.sg

Y. Gong · T. S. Lim · P. Wong
PEC Limited, Jurong, Singapore

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_1,

culling possible colliding pairs or groups. An acceleration data structure called bounding volume hierarchies (BVH) is frequently used to perform the culling (Gottschalk et al. 1996; Tang et al. 2010, 2011). Spatial hierarchical structures such as Octrees (Hunter 1978; Jung and Gupta 1996; Zhou et al. 2010) are also one of those famous acceleration tools. The narrow phase often contains primary interference checks (Govindaraju et al. 2003; Tang et al. 2010, 2011; Ericson 2005) using direct primary tests, separating axes tests, separating plane tests, or proximity tests. As the set of candidate triangles can be very large, traditional narrow phase algorithms are often computationally intensive and time-consuming.

To speed up the algorithms, attempts in developing parallelized collision detection algorithms have been made using multi-core CPUs (Tang et al. 2010) and distributed computers. But none of these platforms could have a stronger parallelization power as modern GPUs do. Researchers such as Fauer and Govindaraju have investigated the potential of relying on GPU platforms to speed up collision detection and have achieved impressive results (Fauer et al. 2008; Govindaraju et al. 2003; Heidelberger et al. 2003). GPUs have a nature parallel-styled design with tremendous computational horsepower and high memory bandwidth to support graphic applications such as shading and rendering (NVIDIA 2010). In order to meet the requirement of more complex general processing and computing, GPUs evolves into General Purpose GPUs (GPGPUs). GPGPUs provide a functional complete set of operations which works on data of arbitrary length. With parallel computing language libraries like CUDA C provided by NVIDIA, programmers are able to allocate GPU memories and run parallel functions called "kernels" in a C/C++ like environment (NVIDIA 2010; Sanders and Kandrot 2011). These new hardware and software designs have brought enormous prosperousness to GPU-enabled computing and processing including collision detection.

Most of the hardware-assisted image-space collision detection methods make use of OpenGL buffers such as depth buffer, stencil buffer, and color buffer (Baciu et al. 1999, Vassilev et al. 2001). But these methods require data read-back, which is often time-consuming due to the asymmetric accelerated graphics port buses in common graphic cards. Moreover, these methods can only be applied in convex objects. Cai et al. (2006) proposed an image-space method using multiple projections to handle convex objects. But the algorithm relies highly on the shape of objects. To achieve correct collision results for complex shapes, the number of projection screens required could be very large.

The *collision detection algorithm* that will be used in this research addresses the following points:

(1) Correctly and efficiently handling arbitrary shapes;
(2) Avoid data read-back in the collision detection process;
(3) Use a single rasterization process to perform collision detection for each component; and
(4) Obtain collision position information during the collision detection process.

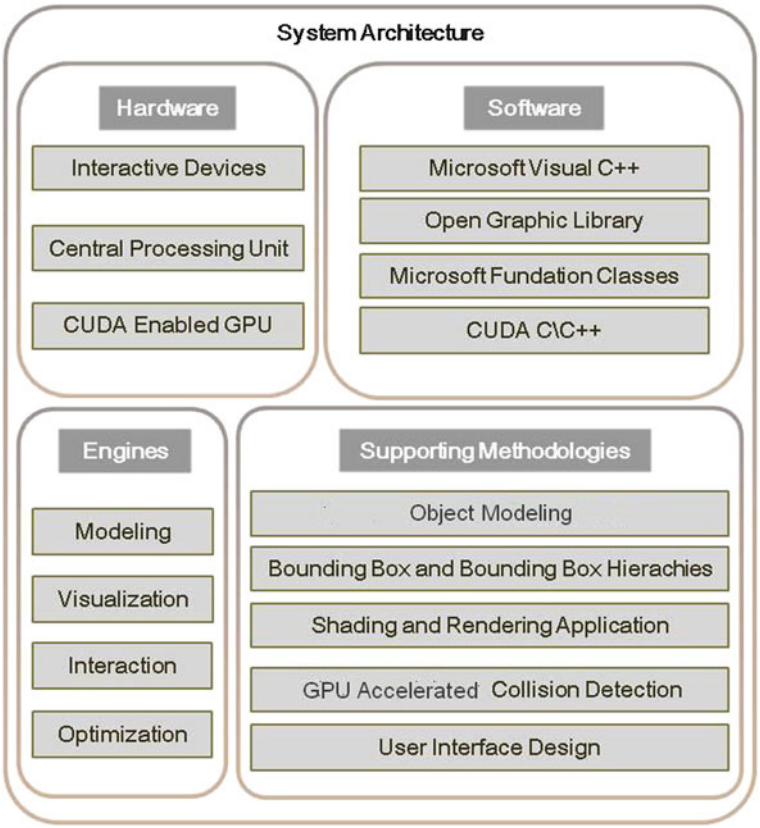


Fig. 1 Simulation system architecture

2 Collision Detection Supported Simulation System Design

Figure 1 illustrates the *overall system structure* for simulation application which consists of three components: *hardware*, *software*, and *methodology*.

2.1 Hardware Component

The fundamental computer hardware used for the system development is a workstation equipped with 1. 86 GHz Intel Core 2 Duo processor E6300 and 4 GB system memory based on windows operating system. The basic simulation operations include *modeling*, *visualization* and *interactions* using central processing unit (CPU). The most challenging aspect to develop a simulator is the *efficient real-time collision detection*. The use of GPU in collision detection can significantly improve the performance of the simulator. In our system design, CUDA enabled GPU

accelerated computing is achieved by using NVIDIA GeForce GTX 560Ti graphic card with 1 GB graphic memory. With 8 multiprocessors and 384 CUDA cores, GeForce GTX 560Ti card provides exceptional performance in real-time scientific rendering and processing. Users are able to launch at most 1,024 threads in a block with a maximum of 65,535 blocks per dimension in a grid.

2.2 Software Component

We develop real-time simulator using Visual C++ compatible on both Windows XP and Windows seven platforms. Open Graphic Library (OpenGL) is used as the Graphic Library for the purposes of model transformation, rasterization and display. Microsoft Foundation Class (MFC) is used for designing the Graphical User Interface (GUI). CUDA is a parallel programming platform supported by NVIDIA graphic cards which enable dramatic increases in computing performance by harnessing the power of the GPU. GPU-based collision test algorithms are developed using CUDA C\C++ in order to optimize the real-time performance.

2.3 Engines and Supporting Methodologies

We have four engines developed taking care of modeling, visualization, interaction, and optimization.

2.3.1 Modeling Engine

Any simulation model is constructed using the *standard geometric modeling techniques*. To optimize the collision detection, rendering, and visualization processes, a hierarchical scene graph is built for the virtual object model from which parameters for users' interaction may be deduced. A scene graph is a tree structure that represents objects or nodes in an order defined by the tree layout. A hierarchical tree structure makes modeling easier because it allows one to construct scene elements based on a hierarchy of objects. By inheriting a coordinate system from their parents, the children automatically follow their parent when the parent is moved. Figure 2 shows a two-party simulation system with the hierarchical tree structure representing each party of object models. Objects are organized into two groups: Party-1 and Party-2. Each of the groups has several components of objects either static or dynamic.

2.3.2 Visualization Engine

3D objects are represented in triangular (polygonal) meshes. The entire vertex coordinates and the vertex indices of triangular faces are extracted from input files

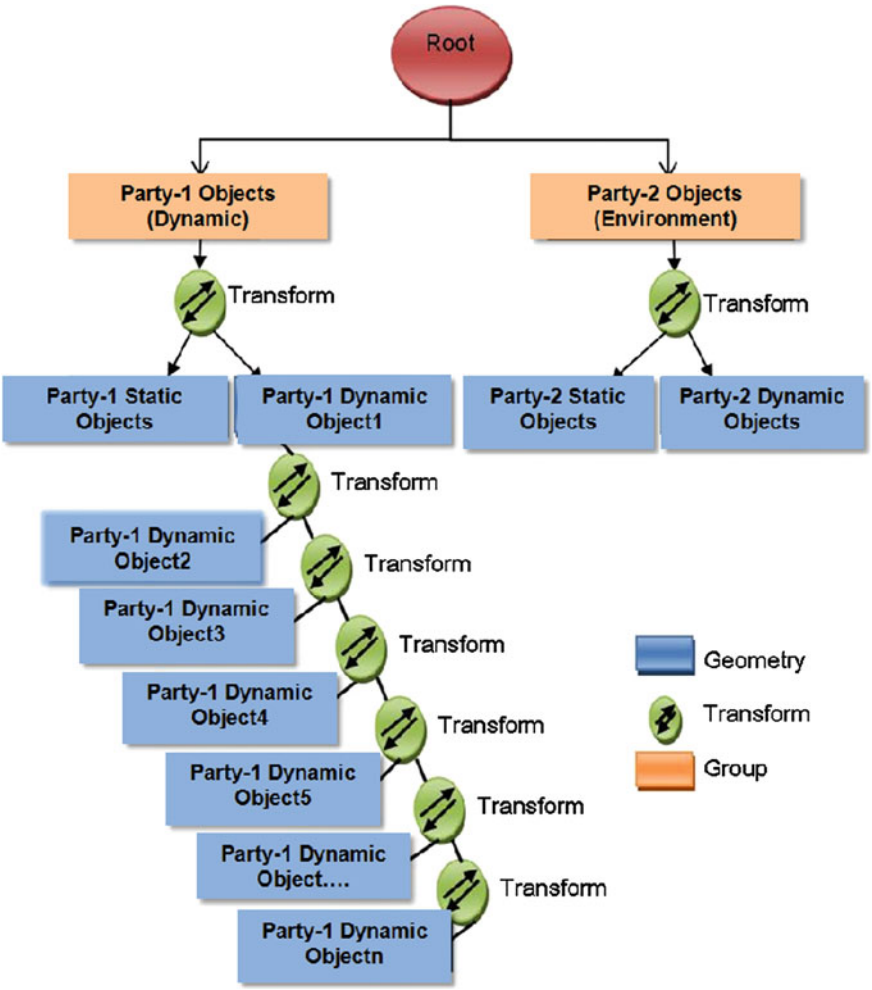


Fig. 2 Illustration of hierarchical tree structure of the two parties

and organized into triangular form to define the geometry of each object. Our system uses the OpenGL to render the 3D object models represented by triangular meshes.

2.3.3 Interaction Engine

Interaction relies on both software and hardware devices. Virtually all interactive devices can be used in the simulation. Selected interactive devices are identified due to the functions required for the simulation purpose. On top of that an interactive GUI is designed using MFC which provides a friendly and convenient manner for users to perform control or manipulation. Typical graphical interactions will include

(1) Translation
(2) Rotation
(3) Scaling
(4) Viewpoint changes (Iso view, front view, side view, etc.)
(5) Undo and redo
(6) Navigation or steering

Along with mouse and keyboard, joystick is also used as input device for our simulation. Generally, joystick device is more suitable for manipulation and transformation providing an interactive experience closer to reality.

2.3.4 Optimization Engine

Parallel features of GPU are explored to speed up the simulator, especially the collision detection engine. The whole collision detection algorithm is implemented in GPU exploiting its capability in rasterization, image compaction, and parallel sorting. Although parallel feature of GPUs is very powerful, performance of the algorithm is still restricted by issues such as huge GPU memory usage caused by high rasterization resolution. To solve this problem, bounding boxes and bounding box hierarchies are introduced to reduce the number of candidate triangles and diminish the raster image size before the rasterization stage. Axis Aligned Bounding Boxes (AABBs) (Ericson 2005) are built for all objects and a bounding volume hierarchy (BVH) is constructed for some object models if needed. As the BVH (Ericson 2005) for the objects contains both AABBs and Oriented Bounding Boxes (OBBs), we name it as a "hybrid BVH."

3 GPU-Enabled Collision Detection

The main idea of the collision detection algorithm is to cast vertical sample rays from a chosen plane and investigate object intersections along the rays to estimate the situation of the scene (Ericson 2005). The process is implemented by rasterization application and the "Sort_Search_Pair" algorithm described in Sect. 3.4.

3.1 Algorithm Overview

The work flow of the GPU collision detection algorithm is as follows:

(1) Update transformation matrices, AABBs and the Party-2 distance array;
(2) Filter Party-2 objects according to the distance map;
(3) Do collision checks between filtered Party-2 objects and the hybrid BVH of the Party-1 objects.

The process of the third step is stated in the following diagram:

Algorithm 1: Pseudo-code of step 3 in the algorithm overview

```
for (each layer of the hybrid BVH)
{
    Launch Kernel1, thd_i work on Object_i in O_r (set of Party-2 objects) and AABB_i in A_r (set of Party-2 AABBs)
    {
        if (AABB_i not intersect with the Party-1 component AABB in the current node)
        {
            Remove Object_i from O_r;
            Remove AABB_i from A_r;
        }
    }
    if (O_r is empty)
    {
        Report "No Collision";
        break for;
    }
    else
    {
        Launch Kernel2, thd_i works on Object_i in O_r and AABB_i in A_r

        {
            if (AABB_i not intersect with the Party-1 component OBB linked to the current node)
            {
                Remove Object_i from O_r;
                Remove AABB_i from A_r;
            }
        }
    }
    if (O_r is empty)
        continue for;
    else
    {
        Store triangles of objects in O_r into T_r;
        Launch Kernel3, thd_i works on Triangle_i in T_r
        {
            if (Triangle_i does not intersect with the Party-1 component OBB linked to the current node)
                Remove Triangle_i from T_r; Compact T_r;
        }
    }
    if (T_r is empty)
        continue for;
    else
    {
        Combine T_r and triangles from Party-1 component mesh data linked with the current node into T_c (set of candidate triangles);
        Pre-rasterize triangles in T_c; (see Section 2.3.4)
        Rasterize triangles in T_c.;
        Launch the "Sort_Search_Pair" algorithm (see Section 2.3.4);
        if (collision detected)
            Report ObjectID and TriangleID;
        else
            continue for;
    }
}
```

3.2 Collision Detection Data Sets

With a big collection of stream processors, the Single Instruction Multiple Data (SIMD) model of GPU (NVIDIA 2010) has significant advantages in processing uniformly represented data. Thus, in this collision detection algorithm, geometric mesh data, bounding boxes, and transformation matrices are stored as GPU styled linear data sets (Fig. 3).

Linear data sets included in the algorithm are:

(1) Party-1 Vertex Set: The coordinates of vertices in the Party-1 object mesh data.
This set is static as the transformations are represented in the matrix streams defined later.
(2) Party 1 AABB Set and Party-1 OBB Set: The corresponding information about AABBs and OBBs maintained for components of the Party-1 objects.
The OBBs will not change once built but the AABBs need to be updated at each rendering time step;
(3) Party-1 Matrix Set: The transformation matrices for Party-1 components dynamically modified from user manipulation.
Vertices in the Party-1 vertex set and OBB information in the Party-1Matrix Set are multiplied by corresponding matrices before use.
(4) Party-2 Vertex Set: The coordinates of vertices in Party-2 object mesh data.
These vertices will also not be changed during the whole lifespan of the simulation.
(5) Party-2 AABB Set and Party-2 Distance Set: AABB representations and Distance value from the center of AABB to the center of Party-1 objects.
These two sets serve as initial sifters for candidate triangles for rasterization.

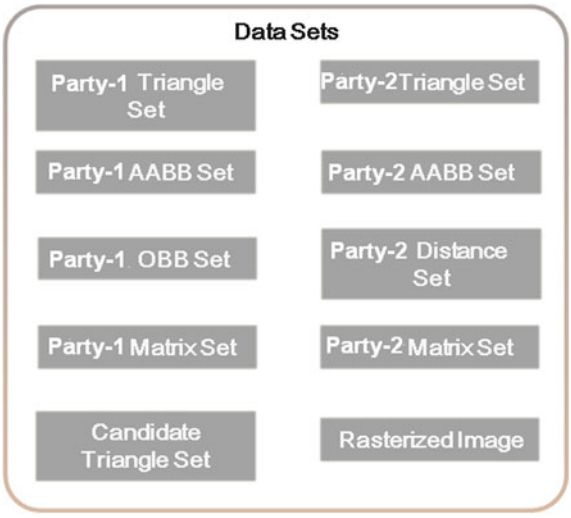


Fig. 3 Linear data sets in the collision detection engine

(6) Party-2 Matrix Set: The transformation matrices for Party-2 objects. The function of this set is similar to the Party-1 Matrix Stream.
(7) Candidate Triangle Set: Vertex references, ID of the objects that the triangles belong to, ID of the triangles in the related objects.
(8) Rasterized Image Set: The image data prepared for rasterization result. The size of the image may change for different Party-1 components.

3.3 Updating the Hybrid BVH

As stated in the algorithm overview, the hybrid BVH is used to cull Party-2 objects and triangles in Party-2 object meshes. The overall structure of the hybrid BVH is a tree with one AABB node and one OBB node at each layer as child nodes of the AABB node in the previous layer. Geometric primitives of Party-1 components are linked to their OBB nodes. The sequence of OBBs in each layer is pre-defined regarding the properties of these Party-1 components. AABBs in layers are built for the set of Party-1 components down from the current layer. The structure of the hybrid BVH is shown in Fig. 4.

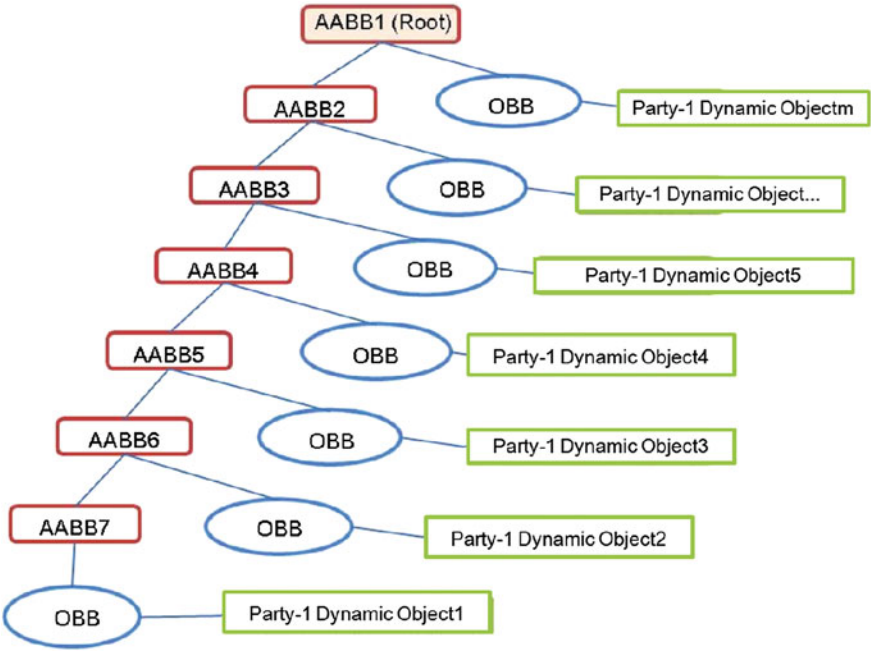


Fig. 4 The hybrid BVH of the crane. *Red blocks* stand for the AABB of the set of all Party-1 components (from object 1 to object m) in lower layers. *Blue blocks* stand for the OBB of the Party-1 component linked to the *blue boxes*

In the BVH updating process, OBBs are first transformed by transformation matrices recording its current direction and location. Then, the system will start the tree traversal to update AABBs via querying for maximum and minimum values of OBB vertices. This can be done in a single query as if we start the search from the bottom in the scene graph shown in Fig. 2.

3.4 Rasterization and the "Sort_Search_Pair" Algorithm

The specially designed rasterization process is the core of the GPU collision detection algorithm. The work flow of it is shown in Fig. 5. Data sets related here are the candidate triangle set and the pixel image set.

First, all the candidate triangles are transformed into the rasterization plane aligned coordinate system. Then a projection matrix is applied on triangle vertices to get their position on the rasterization plane. Depth information is also correspondingly computed.

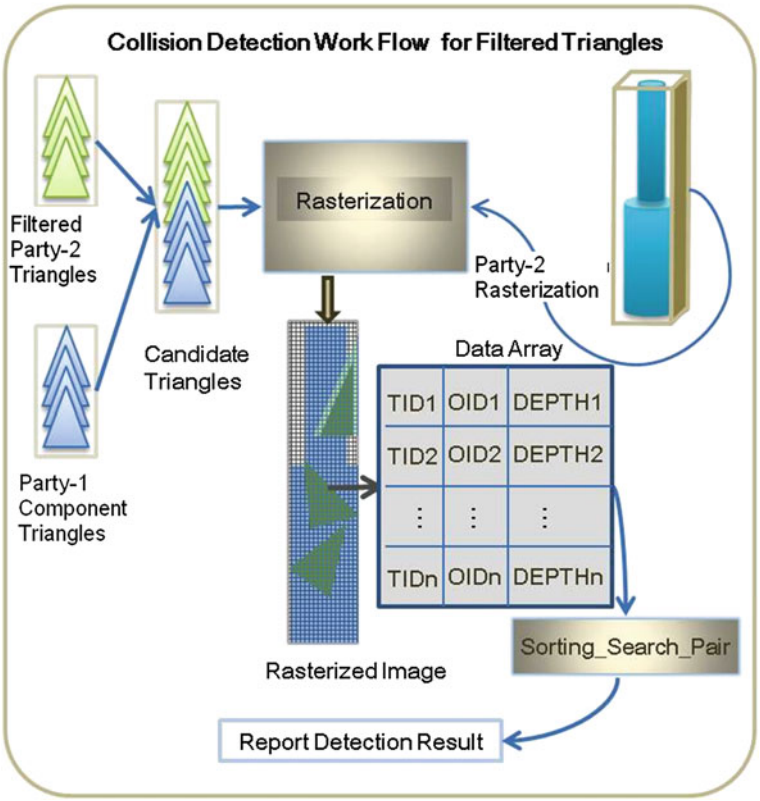


Fig. 5 The rasterization-based collision detection process for filtered Party-2 triangles

In the second step, edge equations of candidate triangles are computed. Functionality of edge equations is to decide the relationships between pixel center points and triangles. Then, for each triangle, a block-shaped GPU parallel kernel is launched for the planar AABB of projected triangle vertices. The algorithm will decide whether pixels in the AABB region are in the overlapped zone of the considered triangle. If the center of a pixel is inside the triangle, the triangle ID, object ID information, and interpolation depth are stored in the corresponding positions in the rasterized image. This stage is referred to as the pixel rasterization stage.

As the numbers of triangles intersecting pixels are not uniform, deciding the size of required memory for the rasterizaed image is a problem. Our solution is to introduce a pre-rasterized process conducted before the pixel rasterization stage to count the amount of triangles overlapping each pixel. Memory needed for rasterization can then be allocated according to the counted result.

The last step is the "Sort_Search_Pair" algorithm which is denoted in Algorithm 2 below:

Algorithm 2: Pseudo-code of the "Sort_Search_Pair" Algorithm

```
Launch Kernel, thd_ij works on pixel_ij in image I (defined in Algorithm 1.1)
{
    Sort the data array d_ij linked to pixel_ij by descending depth value;
    index ←1;
    flag ← 1;
    while (index is no larger than the length of d_ij)
    {
        if (d_ij[index].ObjectID (see in Algorithm 1.1) belongs to Party-1)
            flag ←–flag ;
        else
        {
            if (flag equals -1)
            {
                Output d_ij[index].ObjectID ;
                Output d_ij[index].TriangleID ;
                break for;
            }
        }
        index ← index+1;
    }
    Report "no collision in pixel_ij";
}
```

4 Results

The collision detection part in the interaction engine is not fully implemented yet. Shown in Fig. 6 are some simple collision detection results using the described image-space algorithms for rasterization and collision detection. The blue boxes are Party-1 component and the gray ones denote Party-2 objects. The blue image at the bottom-right corner indicates the rasterized image with the red portion

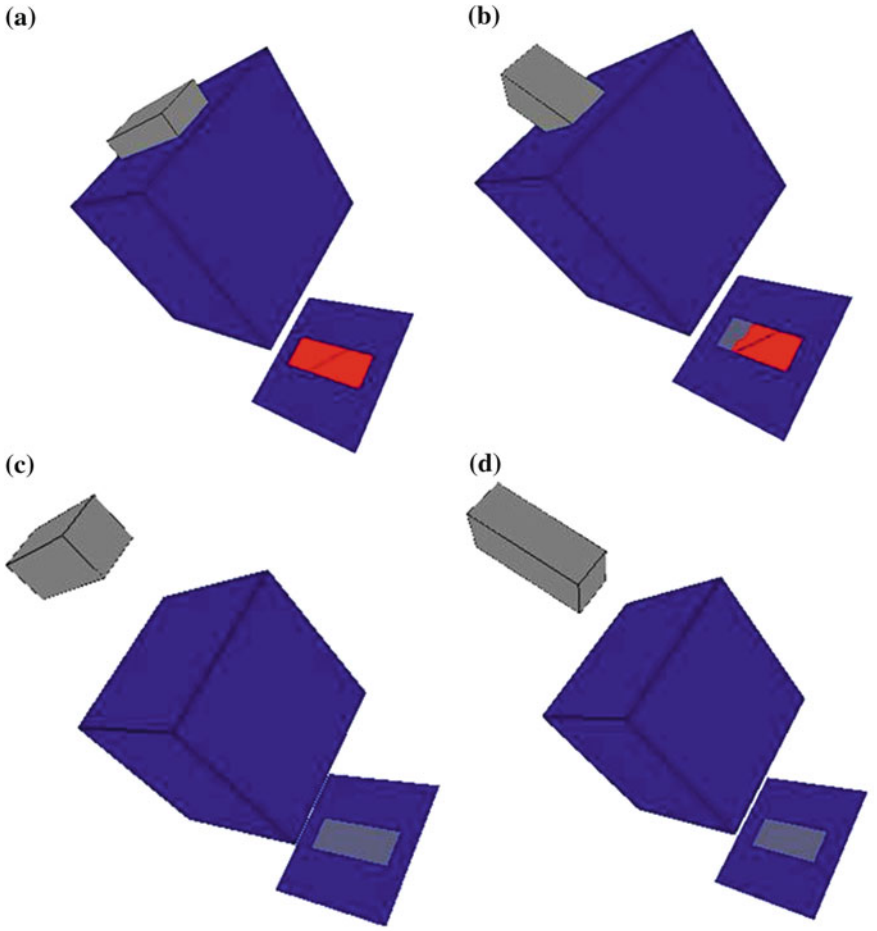


Fig. 6 Simple image rasterization and collision detection results. The *blue box* is one of the Party-1 components and the *gray box* denotes a Party-2 object. The blue image in the bottom-right corner indicates the rasterized image with the red portion referring to collision regions: **a** and (**b**) Collisions detected between the Parties; **c** and (**d**) Situations with no collision

referring to collision regions. Figures 6a, 6b show the cases where the Party-1 component collides with the Party-2 object, while Fig. 6c, 6d shows situations with no collision involved. The blue line in the red region is the overlapping of adjacent triangles in the blue box.

5 Conclusion

In this chapter we investigate the collision detection problem focusing on the use of AABBs technique and the GPU acceleration. Basic issues of simulation modeling, visualization, interaction and user interface are discussed as well. The modeling and hierarchical structure described here enables real-time computing benefiting from the latest GPU hardware acceleration. In particular, image-based algorithm with a hybrid BVH and distance set implemented in GPU is introduced to accelerate collision detection.

There are many applications for collision detection. Robotic path planning is an issue of collision avoidance. Real-time collision detection plays a central role in game, simulation, animation, and computer graphics.

In the future, we will optimize the data structure and the rasterization process to improve the efficiency and numerical accuracy. Implementation of the collision detection engine will be further modified for different industrial needs. Other minor issues such as rasterization plane choosing strategy might be altered to achieve better performance.

References

Baciu G, Wong W, Sun H (1999) RECODE: an image-based collision detection algorithm. J Vis Comput Anim 10(4):181–192

Cai Y, Fan Z, Wan H, Gao S, Lu B, Lim K (2006) Hardware-accelerated collision detection for 3D virtual reality games. Simul Gaming 37(4):476–490

Ericson C (2005) Real-time collision detection. Published in, Elsevier

Fauer F, Sebastien B, Jeremie A, Florent F (2008) Image-based collision detection and response between arbitrary volume objects. Eurographics 2008:155–162

Govindaraju N, Redon S, Lin M, Manocha D (2003) CULLIDE: Interactive collision detection between complex models in large environments using graphics hardware. Graph Hardware 2003:25–32

Gottschalk S, Lin M, Manocha D (1996) OBBTree: A hierarchical structure for rapid interference detection. In Proc. of ACM SIGGRAH'96, pp 171–180

Heidelberger B, Teschner M, Gross M (2003) Real-time volumetric intersections of deformable objects. In Proc Vision Model Vis 2003:461–468

Hunter G (1978) Efficient computation and data structure for graphics. Department of Electrical Engineering and Computer Science, Princeton University, Princeton, NJ Ph.D. dissertation, US

Jung D, Gupta K (1996) Octree-based hierarchical distance maps for collision detection. Int Proc Robot Autom 1996:454–459

NVIDIA (2010) NVIDIA CUDA C Programming Guide. NVIDIA CUDA™

Sanders J, Kandrot E (2011) Cuda by example: an introduction to general-purpose GPU programming. NVIDIA Corporation

Tang M, Manocha D, Tong R (2010) Multi-core collision detection between deformable models using front-based decomposition. Graph Models 72(2):7–23

Tang M, Manocha D, Lin J Tong R (2011) collision streams: fast GPU-based collision detection for deformable models. Symposium on interactive 3D graphics and games, pp 63–70

Vassilev T, Spanlang B, Chrysanthou Y (2001) Fast cloth animation on walking avatars. Comput Graph Forum 20(3):260–267

Zhou M, Gong M, Huang X, Guo B (2010) Data-parallel octrees for surface reconstruction. IEEE Trans Visual and Comput Graphics voi no. pp 669–681

Ink Diffusion Simulation in Chinese Calligraphy Using Navier–Stokes Equation

Yuzhi Ren, Yige Tang, Zhongke Wu and Mingquan Zhou

Abstract This work presents a physically based method using Computational Fluid Dynamics for simulating ink diffusion on paper for traditional Chinese Calligraphy. For the simulation of ink diffusing process, the Lattice Boltzmann Method is employed to solve the Navier–Stokes Equation. And in order to control the flow of the fluid dynamics, a three-layer paper model is designed, which is responsible for the transfer and fixture of the ink particles. For the blurry effect of the boundary of a brushstroke, we devise a multiconcentration ink model, in which the ink of a brushstroke is with different concentration. Finally, a Chinese Calligraphy system is designed, which can produce vivid strokes in real-time.

KeyWords Ink diffusion · Navier–Stokes equation · Lattice Boltzmann model

1 Introduction

In computer graphics, most of the researches on the simulation of painting effects have focused on western paintings, including watercolor, pencil sketching, and hatching (Small 1991; Curtis et al. 1997). Most of the methods work well for the western arts; however, they cannot be directly applied to the Chinese ink painting

Y. Ren · Y. Tang · Z. Wu (✉) · M. Zhou
College of Information Science and Technology, Beijing Normal University, 100875 Beijing, China
e-mail: zwu@bnu.edu.cn

Y. Ren
e-mail: renyuzhi@gmail.com

Y. Tang
e-mail: solidsnake1905@gmail.com

M. Zhou
e-mail: mqzhou@bnu.edu.cn

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_2,

and Calligraphy, due to the special characteristics of Chinese writing tools, which are Chinese brush, ink, and *Xuan* paper. In this chapter, we mainly focused on the ink model and paper model, which together produce the ink diffusion effect.

Guo and Kunii (1991) proposed the first model for the simulation of ink diffusion in Chinese Calligraphy. They assumed that the ink consisted of pigment particles of different sizes, and the ink spreads as a one-dimensional filtering process. Later, a multidimensional diffusion model (Kunii et al. 1995) was proposed, in which the ink diffusion on paper was considered as a two steps procedure: the flow of water on the paper and the motion of the ink particles. This model could simulate precisely the real physical phenomenon, but it required high computational cost for rendering complex strokes; thus it cannot be used for real-time rendering. Kunii et al. (2001) used a partial different equation (PDE) in attempt to better describe the phenomenon, which was essentially Fick's law of diffusion. However, this PDE could only produce blurry strokes instead of the realistic fluid-like patterns. Guo and Kunii (2003) proposed a new algorithm to improve the computational efficiency (Kunii et al. 1995, 2001), by processing only the fronts of the spreading strokes, but this also limited the range of producible effects. Later, a "Pseudo-Brown movement" model (Shi et al. 2003) was used to simulate the impetus of water upon the ink particles, which required high computational cost. The properties of the paper also play a crucial part in the effects of ink diffusion, so some researchers designed different models for the paper (Lee 1999, 2001; Chu and Tai 2005). Lee proposed the paper models of various types to obtain different diffusing results. Laerhoven et al. (2004) described the implementation of a paper model that allowed for real-time creation of watercolor images by simulating the mechanics of pigment and water throughout a three-layer paper. Chu and Tai (2005) proposed a three-layer paper model and used different textures to represent the physical parameters of the paper.

The aforementioned methods mainly focus on the visually plausible diffusion effects, instead of physically accurate results. To obtain physically convincing ink *diffusion patterns*, methods from Computational Fluid Dynamics (CFD) have been exploited. Chu and Tai (2005) adopted the Lattice Boltzmann Equation (LBE) method (He and Luo 1997; Dardis and Mccloskey 1998; Chen et al. 1998; Wei et al. 2004; Chu 2007; Begum and Basit 2008), which models the dynamics of fluid at the particle level.

The main contribution of this research is that a new ink diffusion model based on fluid dynamics has been proposed, and a new and simple three-layer paper model has been designed to simulate the properties of the paper, and together they can produce physically accurate results in real-time. In order to create the blurry of brushstroke boundary, a new multi-concentration ink model is proposed, which can create different blurry effects according to paper.

We present some background on Eastern ink painting and writing in Sect. 2 and describe the Navier–Stokes Equation (NSE) and Lattice Boltzmann Method (LBM) (Succi 2001) in Sect. 3. In Sects. 4 and 5, we present the three-layer paper model and multiconcentration ink model, respectively. Rendering results and conclusions are given in Sect. 6.

2 Physical Properties of Ink Painting

Chinese ink, known as "mo" in Chinese and "sumi" in Japanese, refers to a kind of paint used ubiquitously in the traditional East Asian painting and Calligraphy. Artists use a flexible brush to create expressive lines, shapes, and exploit the interplay of ink and water to produce shades and patterns with spontaneous styles. In order to simulate the diffusion of ink, it's crucial to first explore the physical properties of the ink and the paper during the painting process. In this section, we discuss the properties of the ink and the paper, and the artistic effects related to ink diffusion.

2.1 Physical Characteristics of Chinese Ink and Paper

In general, Chinese ink is composed of glue, soot, and water. The soot is composed of 10–150 nm carbon particles and dissolves readily in water, and those primary particles combine to form relatively bigger clusters (100–300 nm), which represent the physical unit of soot (Swider et al. 2003). When a drop of ink falls on the surface of the absorbent paper, the water begins to spread along the fibers of the paper under the capillary force stemmed from the small spaces in-between the fibers. Some of the ink particles moves with the flowing water, and some remain where they were at the first place. When ink dries, the solid form of Chinese ink is a mixture of carbon particles and glue.

Typical paper consists of fibers organized in random positions and directions: small spaces among the fibers act as small capillary tubes, which can absorb the water and distribute the water to the surrounding places; thus, it produces the effects of diffusion. According to the different diffusion patterns, the Xuan paper can be categorized into three types: untreated Xuan, half-treated Xuan, and treated Xuan. The untreated Xuan is the main medium used in Chinese painting and Calligraphy since it absorbs water easily; treated Xuan paper refers to the Xuan Paper after special treatment with vitriol and gum, and it has the characteristics of no seeping; and the half-treated Xuan's absorptivity is between the above two. In this work, we mainly simulate the diffusion effects on the absorbent Xuan paper.

2.2 Effects of Ink Dispersion

When a brush touches the paper, ink is transferred from the brush onto the paper. Slowly, flowing with the water, carbon particles sink down into the paper and flow along the fibers by the capillary force. Eventually, the water evaporates, leaving the carbon particles fixed on the paper, and the diffusion effects appear. As a result of the dispersing process, the final strokes appear to be bigger than the initial shape.

A remarkable feature of the diffusion effect is a kind of shadow band along the edge of the initial zone, one example is shown in Fig. 1. The complex characteristics

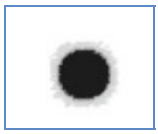


Fig. 1 Diffusion of a drop of ink

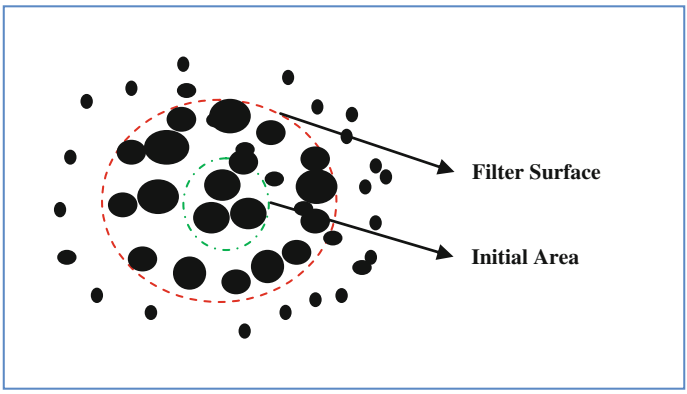


Fig. 2 The filtering effect

of diffusion have a close relationship with the physical interactions between the ink and the paper, which cannot be accurately simulated by conventional graphical functions such as textures, which just produce visually blurred images instead of realistic random diffusion effects. So, the key points of the simulation are how to simulate the flow dynamics of the ink, the spreading and sinking movements of the pigments, and the evaporation of water.

By observation of the real ink diffusing effects (Fig. 1), we can imagine that there is a filter, through which only the carbon particles of the size smaller than the space in-between the fibers can pass; and the larger ones are blocked and remain as they were, as shown in Fig. 2. This is called the *filtering effect*. The area in the red circle denotes the filter surface, and the green one is the initial area where the brush first touches the paper. We implement this filtering effect on the second layer of the three-layer paper model—the flow layer, which is detailed in Sect. 4.1. The size and the density of the particles can be described as the concentration of the carbon particles, which is discussed in Sect. 5.2.

3 Computational Fluid Dynamics

The ink is a kind of colloidal liquid with the viscosity factor and ink diffusion phenomenon can be considered as a typical instance of the diffusion of a colloidal liquid in 2D. The capillary force causes ink to diffuse in the inner structures of the

paper. So, it's physically and practically to use CFD to imitate the flow of ink. We describe the basic NSE and LBM (He and Luo 1997; Shen et al. 1998; Yu et al. 2003) to solve this equation, which is one of the promising methods to solve the NSE.

3.1 The Navier–Stokes Equations

The Navier–Strokes equations (Stam 1999) is obtained by imposing the constraints that the fluids conserve both the mass and the momentum. Although factors such as temperature and density may influence the flow, we only use two parameters to describe the flow, a velocity field u and a pressure field p, which are functions of both time and spatial positions. Here, we denote the fluid as x, u both at two dimensions $\{x, y\}$. Given the velocity and the pressure at the initial time$t = 0$, the evolution of these quantities over time is given by NSE.

$$\nabla \cdot u = 0 \tag{1}$$

$$\frac{\partial u}{\partial t} = -(u \cdot \nabla) - \frac{1}{\rho}\nabla p + v\nabla^2 u + f \tag{2}$$

where $\nabla = (\partial/\partial x, \partial/\partial y)$, "$\cdot$" denotes the dot product, and the shorthand notation $\nabla^2 = \nabla \cdot \nabla; v$ and ρ are the kinematic viscosity and the density of the fluid respectively, and f is the external force.

The pressure and the velocity fields in the NSE are related. A single equation for the velocity can be obtained by combining Eqs. (1) and (2). The Helmholtz–Hodge Decomposition can be used that any vector field w can uniquely be decomposed into the following form:

$$w = u + \nabla q \tag{3}$$

where u has zero divergence: $\nabla \cdot u = 0$ and q are a scalar field. Any vector field is the sum of a mass conserving field and a gradient field. This result can define an operator P which projects any vector field w onto its divergence free part $u = Pw$. The operator is in fact defined implicitly by multiplying both sides of Eq. (3) by "∇":

$$\nabla \cdot w = \nabla^2 q \tag{4}$$

This is a Poisson equation for the scalar field q with the Neumann boundary condition for the scalar field q with boundary condition $\partial q/\partial n = 0$. A solution of this equation is used to compute the projection u:

$$u = Pw = w - \nabla q \tag{5}$$

If we apply this projection operator on both sides of Eq. (2), we obtain a single equation for the velocity:

$$\frac{\partial u}{\partial t} = P\left(-(u \cdot \nabla)u + v\nabla^2 u + f\right) \tag{6}$$

where $Pu = u$ and $P\nabla p = 0$. This is the fundamental equation from which we will develop a stable fluid solver.

3.2 Method of Solution

Because the NSE is hard to solve, various methods have been proposed. The traditional methods (Chorin 1968) solve for the macroscopic variables, such as velocity and density, which is complex. Unlike the traditional ones, the Lattice Boltzmann Method (LBM) is based on the microscopic kinetic equation for the particle distribution function. The LBM has become an attractive method to the conventional computational fluid dynamics method for solving NSE. Because we are going to use to simulate the ink diffusion, we would like to introduce a standard square lattice model for 2D flow called D2Q9, as shown in Fig. 3, which are just like ink diffusion on paper. The physical space is divided into a regular lattice and the velocity space is discretized into a finite set of velocities c_α, the Boltzmann equation can be discretized as

$$f_\alpha(x_i + c_\alpha \Delta t, t + \Delta t) = \left(1 - \frac{1}{\tau}\right) f_\alpha(x_i, t) + \frac{1}{\tau}\left(f_\alpha^{\mathrm{eq}}(x_i, t)\right) \tag{7}$$

where Δt and $c_\alpha \Delta t$ are time step and space increments, respectively, f_α is the single-particle velocity distribution function along the α_{th} direction; f_α^{eq} is the equilibrium distribution function, and τ the single relaxation time.

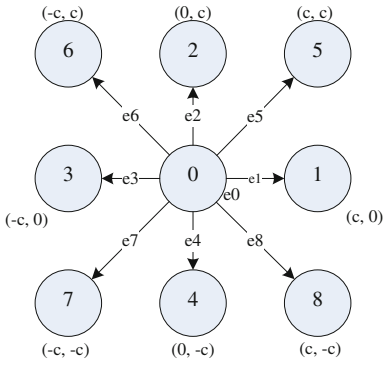


Fig. 3 D2Q9

There are different types of lattice for LBM. For simplicity and without loss of generality, we consider the two-dimensional square lattice with nine velocities as shown in Fig. 3, the D2Q9 model:

$$c_\alpha = ce_\alpha = \begin{cases} (0,0) & \alpha = 0, \\ c\left[\cos\left(\frac{\alpha-1}{2}\pi\right), \sin\left(\frac{\alpha-1}{2}\pi\right)\right] & \alpha = 1,2,3,4 \\ \sqrt{2}c\left[\cos\left(\frac{\alpha-5}{2}\pi + \frac{\pi}{4}\right), \sin\left(\frac{\alpha-5}{2}\pi + \frac{\pi}{4}\right)\right] & \alpha = 5,6,7,8 \end{cases} \tag{8}$$

Here, we set $c = 1$, $\rho = 1$. For thermal fluids, the equilibrium distribution function for D2Q9 model is given by

$$f_\alpha^{\mathrm{eq}} = \rho w_\alpha \left[1 + \frac{3}{c^2} c_\alpha \times u + \frac{9}{2c^4}(c_\alpha \times u)^2 - \frac{3}{2c^2} u \times u\right] \tag{9}$$

where $w_o = \frac{4}{9}, w_1 = w_2 = w_3 = w_4 = \frac{1}{9}$, and $w_5 = w_6 = w_7 = w_8 = \frac{1}{36}$. The macroscopic density ρ and velocity u are related to the distribution function by using the Chapman-Enskog expansion; Eq. (7) can recover the NSE to the second order of accuracy, with the kinematic viscosity given by

$$\rho = \sum_{\alpha=0}^{8} f_\alpha \quad \text{and} \quad \rho u = \sum_{\alpha=0}^{8} f_\alpha c_\alpha \tag{10}$$

$$v = \frac{(\tau - 0.5)\mathrm{c}^2 \Delta t}{3} \tag{11}$$

In the standard LBM, Eq. (7) has a two-step process: the streaming step and the collision step, as follows:

(1) Streaming step:

$$f_\alpha(x_i + c_\alpha \Delta t, t + \Delta t) = \widetilde{f}_\alpha(x_i, t) \tag{12}$$

(2) Collision step:

$$\widetilde{f}_\alpha(x_i, t) = f_\alpha(x_i, t) - \frac{f_\alpha(x_I, t) - f_\alpha^{\mathrm{eq}}(x_i, t)}{\tau} \tag{13}$$

4 Three-Layer Paper Model

The movement of the ink in the paper is determined by the fluid dynamics and the interactions between the ink and the paper as well. Here, we adopt the LBM to simulate the ink movement. However, unlike the fluid dynamics simulation, the ink should stop dispersing after the water is absorbed by the paper and finally evaporate. Therefore, in order to effectively control the flow range of the ink fluid, a three-layer paper model similar to (Laerhoven et al. 2004; Chu and Tai 2005) is

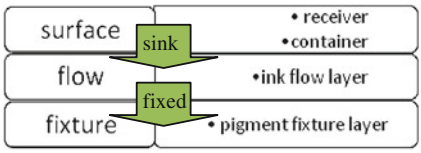


Fig. 4 Three-layer paper model

used. In this section, we will talk about the three-layer paper model, the transfer of ink, the pigment fixture, and the ink evaporation.

4.1 Paper Model

The three-layer paper model consists of a surface layer, a flow layer and a fixture layer, as shown in Fig. 4. The top surface layer acts like a receiver and a container, which receives the ink from the brush and conserves the ink; the middle flow layer only gets water from the surface layer, which is responsible for ink flowing along the fibers of the paper; with the ink flowing, the smaller ink particles sink down into the bottom fixture layer, and finally it shows the result of diffusion. In this section, we will discuss the transitions of the ink through each of the three layers.

4.2 Transfer of Ink in Surface Layer

In our real painting process, the amount of ink deposited on paper has a close relationship with the results. The amount of ink deposited onto the surface is determined by the amount of ink contained in the brush hand, the size of brush footprint on paper surface. Ink is supplied from the surface layer to the flow layer according to the capacity of the paper fibers. The surface layer stores excess ink not yet absorbed by the flow layer.

To model the capacity of the paper, to model the different water, the amount of water supplied to flow layer is clamp $(s, 0, \pi - \rho)$, where ρ represents the density of water in the pixel in the flow layer, which could be calculated from Eq. (10), π represents the capacity of the paper fiber and clamp (s, min, max) returns the clamped value of s against the limits of min and max. What's more, each pixel in the footprint is masked by the value max $\left(1 - \frac{\rho}{\pi}, m\right)$, where m is base mask value, for example, the paper texture (Chu and Tai 2005).

When ink drops on the surface, the first layer is like a reservoir. We represent the ink concentration on the surface layer as p_s. At the same time, clamp $(s, 0, \pi - \rho)$ amount of ink sinks down into the flow layer, whose main task is to simulate the capillary attraction in the paper. So, the concentration of ink in the flow layer is $\frac{p_f \cdot \rho + p_s \cdot \text{clamp}(s,0,\pi-\rho)}{\rho + \text{clamp}(s,0,\pi-\rho)}$. According to the parameters: p_f, p_s denote the concentration of ink in the flow layer and in the surface layer at the last frame, respectively.

4.3 Fixture of the Ink in Fixture Layer

The final results are represented on the third layer. To model the pigment in the third layer realistically, three rules are designed: (1) the transfer rate, the percentage of ink fixed on the third layer compared to the total amount in the same position, is higher when the glue is more concentrated; (2) the transfer rate is higher as the strokes become drier; (3) all the carbon particles are settled when a stroke dries.

We devise a simple pigment fixture algorithm that satisfies the above rules. ρ', the water density (amount) in the flow layer in the last frame; μ, ξ, denote parameters for modulating the fixture rate by dryness and glue, respectively. According to other parameters, g is the glue concentration in the flow layer; and η is the base fixture rate.

In Algorithm 1, rows 2–4 assure that if the stroke becomes drier, that is, loss >0, then fix factor is higher. The row 5 makes sure that the temp is in the range 0–1, while, the temp is increasing with μ, ξ. In row 6, in order to produce boundary roughening and the color is darker than paper, we adopt the smoothstep function and max function, separately. Rows 7 and 8 update the data in the surface and flow layer respectively.

Algorithm 1

```
1  void fixture(p_f, p_x, ρ, ρ'){
2        loss ← ρ' − ρ
3      if (loss>0)   fixfactor ← wloss / ρ'
4      else          fixfactor ← 0;

5        temp ← clamp(μ + ξ * g, 0,1)
6        fixfactor ← max (fixfactor * (1 − smoothstep(0, temp, ρ)), η)
7        p_x ← p_x + fixfactor * p_f
8        p_f ← p_f − fixfactor * p_f
9  }
```

4.4 Water Evaporation from Flow Layer

In real writing process, when water evaporates from the paper, the pigment is left and the stroke appears. Every time evaporation happens, the pigment in the fixture darkens. So, in our imitating, evaporation is essential and important. We model this by having different evaporation rates for the boundary sites and the rest of the wet sites because of the different contact area rate of water. We do this by reducing

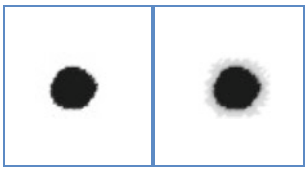


Fig. 5 Generated images with diffused ink drop. **a** With this new multiconcentration ink model and **b** with the single concentration ink model

the water density ρ at a rate of *evaporation_rate* as we update the ρ using Eq. (9), while it will increase the pigment in the fixture layer. This process is similar in (Chu and Tai 2005), except that we do not reduce those f_i' that bounce back during streaming step.

In Algorithm 2, rows 2 and 3 obtain the evaporation rate. Rows 4 and 5 update the data in the fixture layer and flow layer respectively.

Algorithm 2

```
1  Void Evaporate (p, b, a,){
2      if (p== boundary sites)    evaporation_rate ← b;
3      else                       evaporation_rate ← a;

4      p_x+←evaporation_rate*ρ*pf;
5      p_f *←(1-evaporation_rate);
6  }
```

5 Multiconcentration Ink Model

In the real calligraphy, the boundary of a brushstroke is less darker than the initial part and the pigment isn't changing sharply, which calls the blur effect. With the traditional single ink model, the pigment of the diffused image change sharply, which could not create the blurry effect, shown in Fig. 5a. In order to simulate the ink physically and simply, we devise a new ink model for this system.

5.1 Ink Model

Although the real ink in the same brushstroke is with the single concentration and the same quantity, in this new ink model, we assume that the ink of a brushstroke is with different concentration and different quantity. Generally speaking, when a

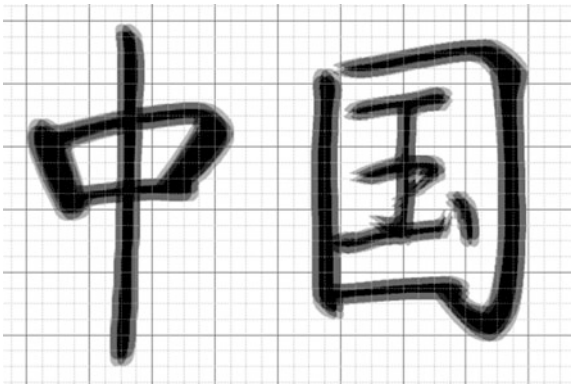


Fig. 6 The ink distribution in the surface layer

new brushstroke is generated, the surface layer gets concentration and quantity of ink from the brush, the middle part of the brushstroke have a higher concentration and more ink than the boundary parts. In Fig. 6, it shows the distribution of ink in the surface layer, in which the black part with more ink and has a higher concentration.

What's more, the diffusion effect is connected directly to the paper's absorbability, the wetness of the brush, the viscosity and concentration of ink. We will adopt a new algorithm to preprocess the ink on the surface layer.

5.2 Different Ink Concentrations

The blurry effect has something to the paper absorptivity, the viscosity and concentration of ink. In this paper, we use the viscosity, v, which we could get from Eq. (10). From observations and researches, the higher the paper absorptivity is, the blurry effect can be more obvious. While the higher the ink viscosity is, the blurry effect can be less obvious. In this research, we ignore the absorptivity and only take the viscosity into account. We devise an algorithm to process ink before diffusion begins.

In Algorithm 3, row 2 makes sure that the whole algorithm processes the data on the boundary. Rows 4 and 5 assure that the temp is related to the absorptivity to the paper. Rows 5–12 renew the concentration and quantity of ink in the boundary in the surface layer.

In Fig. 7, the images show the different diffusion effect due to the viscosity. (a) diffused with the traditional single ink model; from (b)–(d), the images all diffused with the multiconcentration ink model, while the viscosity is increasing.

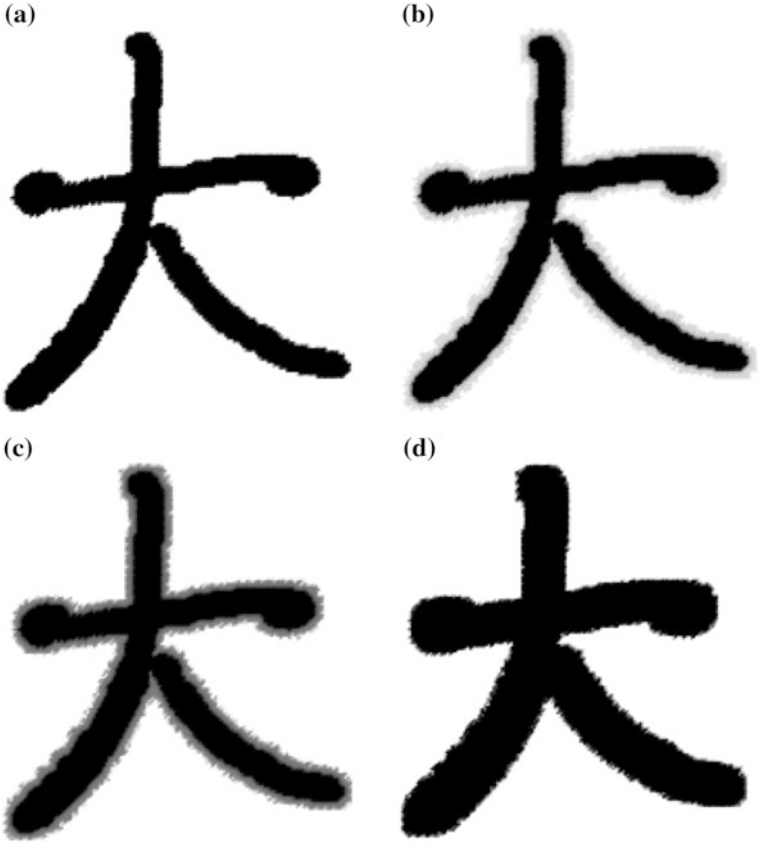


Fig. 7 Images of a Chinese word of diffused ink. **a** With the traditional single concentration model and **b–d** with the multiconcentration ink model while the viscosity is increasing

Algorithm 3

```
1   void process_ink ( p_s, v, l_s,) {
2       if (p !=boundary_sites)    return;
3       else      temp ← absorptivity
4       if (temp<1)   temp ←1;

5       for  i=0 to 8
6            new_p←p+e[i]*temp;
7            if  new_p = out of boundary{
8               p_s[new_p] ← v * p_s[p]
9               l_s[new_p] ← v * l_s[p]
10           }
11      end for
12  }
```

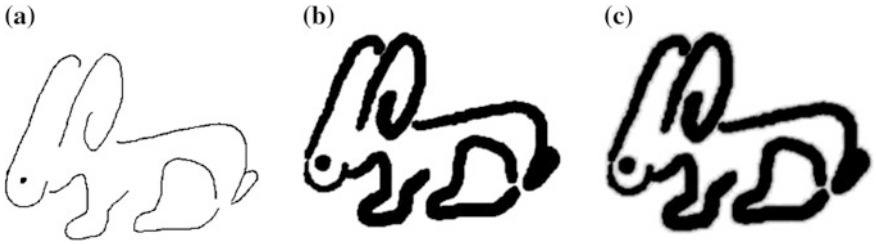


Fig. 8 Rabbits **a** without ink diffusion and without pressure, **b** diffusion with the single concentration ink model, and **c** diffusion with the multiconcentration ink model

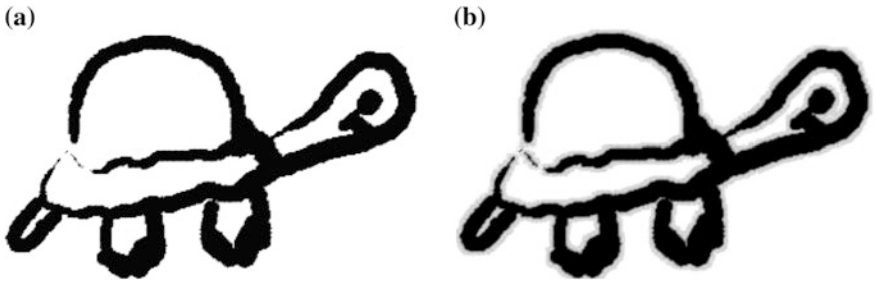


Fig. 9 Turtles **a** diffusion with the traditional single ink model and **b** diffusion with multiconcentration ink model

6 Results and Conclusion

6.1 Results

All the results are rendered on a desktop PC with a *3.07 GHz IntelCore i7 CPU* with 4 GB of RAM CPU, which is equipped with a *NVIDIA GeForce GTX 570 GPU* with 1280 MB of dedicated video memory. The *Wacom Cintiq 21UX Pen Display* device is used to obtain the information of the pressure and the positions during the drawing interaction, and used as the display monitor as well.

We have designed an interaction system for the purpose of Chinese Calligraphy education, and focus on the simulation of ink diffusion effect, which is essential for showing the vivid brushstrokes of Chinese Calligraphy. The rendering results are computed in real-time as a resolution of 512×512, showing the differences between the results without and with ink diffusion effect; and more results are shown in Figs. 8, 9, and 10.

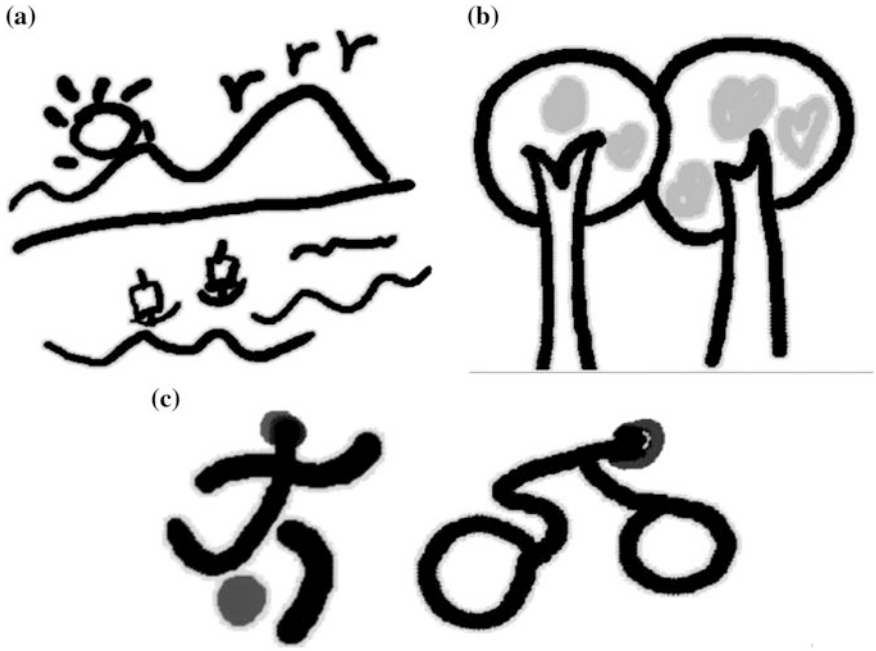


Fig. 10 **a** Simple Chinese landscape Painting, **b** tree, and **c** sports

6.2 Contribution

In this chapter, a novel approach is presented for simulating ink diffusion on paper. First, the flow of the ink is realized based on the fluid dynamics, and the Lattice Boltzmann method is used to solve the Navier–Stokes equations; therefore, it can produce physically realistic diffusing results. Second, a three-layer paper model has been designed to control the flow. The surface layer is responsible for receiving ink from brush; the ink only flows in the flow layer; as the ink evaporates from the flow layer, the carbon particles gather in the fixture layer. Thirdly, in order to produce blurry effects, a multiconcentration ink model, which is based on the ink viscosity and paper absorptivity, is devised to control the blurry results. Finally, a Chinese Calligraphy system is implemented based on the above models, which produce vivid strokes in real-time.

Besides ink diffusion, the Chinese brush also plays an important role in the artistic effects of the Chinese Calligraphy. So our future work is to design a more accurate physically based model for the brush, which can produce more complicated, vivid, and realistic strokes.

References

Begum R, Basit MA (2008) Lattice Boltzmann method and its applications to fluid flow problems. Eur J Sci Res 22(2):216–231

Chen S, Doolen GD (1998) Lattice Boltzmann method for fluid flows. Annu Rev Fluid Mech 30(1):329–364

Chorin AJ (1968) Numerical solution of the Navier–Stokes equations. Math Comput 22(104):745–762

Chu SH (2007) Making digital painting organic. Dissertation, Hong Kong University of Science and Technology, People's Republic of China

Chu NSH, Tai CL (2005) MoXi: real-time ink dispersion in absorbent paper. ACM Trans Graph 24(3):504–511. doi:10.1145/1073204.1073221

Curtis CJ, Anderson SE et al (1997) Computer-generated watercolor. In: Proceedings of the 24th annual conference on Computer graphics and interactive techniques. ACM Press, pp. 421–430. doi: 10.1145/258734.258896

Dardis O, Mccloskey J (1998) Lattice Boltzmann scheme with real numbered solid density for the simulation of flow in porous media. Phys Rev E Lett 57(14):4834–4837

Guo Q, Kunii TL (1991) Modeling the diffuse painting of sumie. In: IFIP modeling in computer graphics, pp 329–338

Guo Q, Kunii TL (2003) Nijimi rendering algorithm for creating quality black ink paintings. In: Proceedings of computer graphics international 2003, pp 152–159

He X, Luo LS (1997) Lattice Boltzmann model for the incompressible Navier–Stokes equation. J Stat Phys 88(3):927–944

Kunii TL, Nosovskij GV et al (1995) A diffusion model for computer animation of diffuse ink painting. In: Proceedings of the computer animation. IEEE Computer Society, p 98

Kunii TL, Nosovskij GV et al (2001) Two-dimensional diffusion model for diffuse ink painting. Int J Shape Model 7(1):45–58

Laerhoven T, Liesenborgs J, Reeth F (2004) Real-time watercolor painting on a distributed paper model. In: Proceedings of Computer Graphics International 2004, pp 640–643

Lee J (1999) Simulating oriental black-ink painting. IEEE Comput Graph Appl 19(3):74–81

Lee J (2001) Diffusion rendering of black ink paintings using new paper and ink models. Comput Graph 25(2):14

Shi Y, Sun J et al (2003) Graphical simulation algorithm for Chinese ink wash drawing by particle system. J Comput Aided Des Comput Graph 15(6):667–672

Small D (1991) Modeling watercolor by simulating diffusion, pigment and paper fibers. SPIE Proc 1991:140–146

Stam J (1999) Stable fluids. In: Proceedings of the 26th annual conference on Computer graphics and interactive techniques, ACM Press, pp 121–128

Succi S (2001) The lattice Boltzmann equation for fluid dynamics and beyond. Oxford University Press, Oxford

Swider JR, Hackley VA et al (2003) Characterization of Chinese ink in size and surface. J Cult Herit 4(3):175–186

Wei X, Zhao Y et al (2004) Lattice-based flow field modeling. IEEE Trans Visual Comput Graph 10(6):719–729

Yu D, Mei R et al (2003) Viscous flow computations with the method of lattice Boltzmann equation. Prog Aerosp Sci 39(5):329–367

Cloth Simulation and Virtual Try-on with Kinect Based on Human Body Adaptation

Yuzhe Zhang, Jianmin Zheng and Nadia Magnenat-Thalmann

Abstract This chapter describes a virtual try-on system with Kinect, which includes three components: data extraction from Kinect, animated body adaptation, garment prepositioning and simulation. With Kinect, some measurements and motion data of the customer can be extracted from captured RGB, depth and motion information. An anthropometry-based method is proposed to generate a customized 3D body from a template model. The method involves three processes. First, a statistical analysis method is proposed to estimate the anthropometric measurements based on partial information extracted from Kinect. Second, a constrained Laplacian-based deformation algorithm is presented to deform the template model to match the obtained anthropometric measurements. Third, a shape refinement method based on the contours is presented. The customized model is then animated by the recorded motion data from Kinect. Meanwhile, a scheme for automatic prepositioning the 2D patterns of garments onto the customized body is proposed. After the step of topological stitching and simulation, the garment is dressed on virtual model of customer finally. With the proposed framework, customers can try-on designed garments in an easy and convenient way. The experiments demonstrate that our try-on system can generate accurate results.

Keywords Cloth simulation · Kinect · Human body adaptation

1 Introduction

Virtual try-on is a technique that allows users to *virtually dress garments onto digital 3D body.* It has many applications. For example, in the scenario of online cloth shopping, a customer picks clothes in the category and the result of dressing

Y. Zhang (✉) · J. Zheng · N. Magnenat-Thalmann
Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore
e-mail: YZHANG3@ntu.edu.sg

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_3,

the clothes on his or her body will be displayed on the screen to check whether the clothes fit or not. This process involves many technical components such as 3D body modeling, adaptation, clothes preposition, and simulation. Also a simple interface and realistic dressing results are expected.

Extensive research has been conducted on the modeling of human body and the simulation of virtual clothes. Substantial progress has been achieved in these areas. The development of virtual human modeling and virtual cloth simulation makes the virtual human and garment realistic and impressive (Volino and Magnenat-Thalmann 2000; Magnenat-Thalmann and Volino 2005; Magnenat-Thalmann 2010). However, these techniques are in general too complicated or need sophisticated equipment, which make the virtual try-on system still far away from practical use. Considering the popularity and cheap price of Kinect, this chapter studies technologies for virtual try-on with Kinect.

1.1 Related Work

Cordier et al. (2003) proposed the concept of made-to-measure online clothing store. This online try-on system includes two databases for body and garments, respectively, and supports interactive manipulators for customers to adjust 3D mannequin by controlling measurements. 3D garment is resized to provide online fitting. It integrates the techniques of virtual human and cloth simulation to achieve the goal of dressing for an individual customer. However, there are some drawbacks in the system. First, the system needs many indirect and complicated interactions for model generation. Second, the body model which is created through only eight measurements is not very accurate. Third, the system generates the initial garment on a generic model in advance and then performs simple deformation according to the shape of the customer's model. The 3D garment may not follow the initial designed 2D patterns of garments.

Founded in 2012, the company Bodymetrics (http://www.bodymetrics.com/) investigates a system for virtual try-on that uses the same Kinect's cameras technology to reconstruct the user's body and uses it to virtually try and order clothing that fits online. However, the try-on result is statistic so that users cannot have a sense of dynamic effect of cloth with regards to their movement.

Body model generation is a research component needed in virtual try-on. Many methods have been proposed, which are divided into four groups: scanner-based methods, photo-based approaches, example-based, and anthropometry-based methods. Scanner-based methods usually reconstruct surface of body model by meshing point clouds from 3D Laser scans, which are accurate but expensive and time consuming. Photo-based methods, such as (Barron and Kakadiaris 2000) and (Seo et al. 2006), reconstruct body model based on multiple images from different views or a single image. In this kind of methods, the 3D surface results are inaccurate. And pose registration is difficult, which can be solved easily by Kinect. Example-based methods, such as (Allen et al. 2003) and (Seo and Magnenat-Thalmann 2004),

generate a quite precise and high qualify model, however, it is inconvenient to do local modification and modify models according to user's intention. The approach of parameterizing the human body according to measurements can generate body model in real time and is more suitable for digital fashion applications since anthropometry is also widely used in fashion industry. After Grosso (Grosso et al 1987) first introduced anthropometry-based modeling, many researchers (Shen et al. 1994; Maïm et al. 2009; Kasap and Magnenat-Thalmann 2010, 2011) provided user interactive manipulators to create human models through anthropometric measurements. However, most existing methods treated the anthropometric parameters independently. As a consequence, unrealistic model may result. Thalmann (Magnenat-Thalmann et al. 2004) combined example-based and anthropometry-based method to generate body model satisfying multiple measures, however, it is not a real-time method and lacking further refinement.

Virtual try-on relates the clothes with 3D body. To dress 3D body with garments defined by 2D patterns, a prepositioning step is needed. The traditional manual placement requires intensive interaction for designers to position patterns one by one. Research works have been studied to achieve semi-automatic prepositioning. For example, Groß et al. (Fuhrmann et al. 2003; Groß et al. 2003) place and arrange patterns on an extendable expanded cone surface first and then map this cone surface to trunk or limbs of body which is also approximated by cone surface. Fontana et al. (2006, 2008) set up a hierarchical structure to store the garment information in different levels. During prepositioning, they use several developable surfaces to approximate body model first and map patterns to body according to embedded knowledge. Thanh and Gagalowicz (2005, 2009) provide a tool to map patterns onto front and back 2D silhouette of model. All these approaches try to link patterns with the body model through connectivity of patterns. Based on the summary of three methods, we infer that semantic-based representation to encode linking information between pattern and body is required and the method to encode the position information to transform patterns of garment to different body model automatically is missing.

There is a huge amount of research focused on cloth simulation. Different schemes such as finite elements and particle systems are proposed to produce the dynamic behavior of clothes. There are also a lot of works on numerical methods to improve the efficiency of the simulation or collision detection and response which are to detect the contacts between regions of the clothes with the body and other parts of the clothes to simulate reaction and friction force. A review of cloth simulation can be found in (Volino and Magnenat-Thalmann 2000; Magnenat-Thalmann and Volino 2005; Magnenat-Thalmann 2010).

1.2 Our Work

Our work is inspired by Kinect and its applications (Izadi et al. 2011). Kinect was launched by Microsoft in 2010 and SDK for Window 7 was released in June, 2011. Since Kinect is a controller-free device and can track the movement of a full body

in real time, Kinect is changing the way in which people interact with computer. Our goal is to develop efficient techniques with Kinect for automatic generation of virtual mannequin and dressing garment on it.

In this research, we present a Kinect-based virtual try-on system. Figure 1 illustrates the technical components of the proposed system, which are mainly data collection, body generation, and garment dressing. In the stage of data collection, the customer only has to show a T-pose in the front view and a straight posture in the side view in front of the Kinect, then move or pose the body for cloth dressing. The camera and sensor of Kinect will capture the movement of the customer. The system will analyze, filter, and record the data stream to extract useful information. There are three kinds of information extracted from the data captured by Kinect. The first includes some measurements of the customer body such as the breadth of shoulder, waist and hip, and the length of limbs. The second is the contours of the customer in the front and side views. The third is the motion data of the customer. The measurements and the contours of the customer are used to deform and adapt a template body. The initial static mannequin of the customer will then be animated by the motion data captured by Kinect. Finally, the customer can select a garment and the try-on system will automatically position the patterns of the picked garment

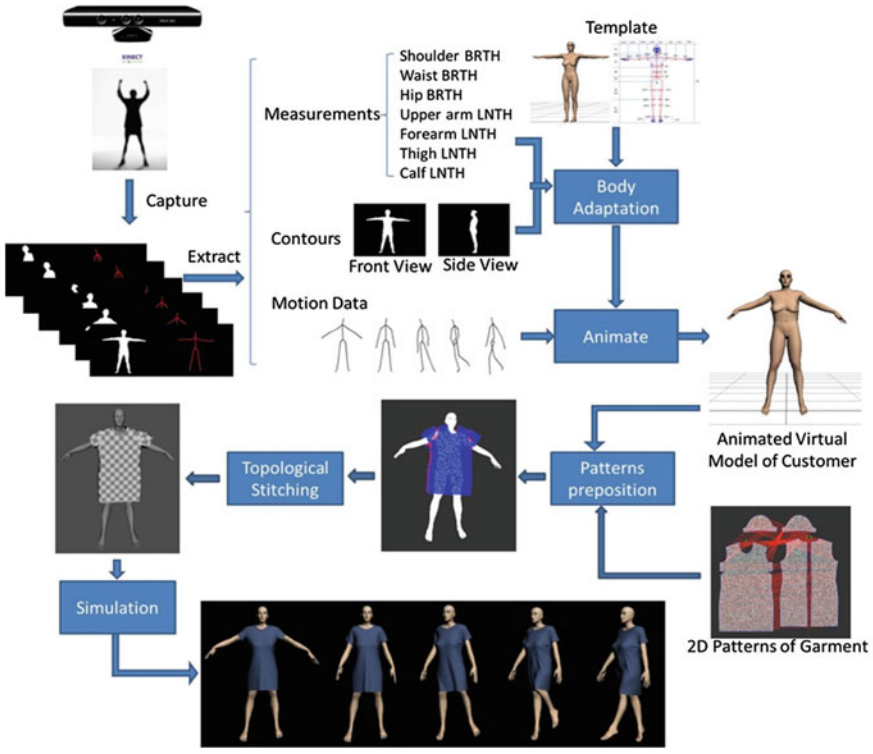


Fig. 1 Workflow of the proposed virtual try-on

around the mannequin of the customer and the dynamic simulation of the virtual body dressed with the picked garment will be displayed.

The main contributions of the paper include:

(1) We present a virtual try-on system with Kinect, which allows users to virtually try-on garments on their own digital body.
(2) We present a body adaptation algorithm based on human anthropometry, which involves statistical analysis of anthropometric measurements, measurement driven surface deformation, and shape refinement based on contours. Compared to the existing anthropometry-based approaches, our approach considers the correlation among the measurements and thus generates more accurate models.
(3) We present an automatic dressing garment method based on prepositioning, topological stitching, and simulation.

This chapter is organized as follows. Section 2 explains data collection from Kinect. The algorithms for measurement prediction, anthropometry-based body deformation, and contour-based shape refinement are presented in Sect. 3. Garment prepositioning and simulation are presented in Sect. 4. Section 5 gives some experiment results and Sect. 6 concludes the research.

2 Kinect-Based Data Collection

Kinect can capture three types of data: RGB frames, depth frames, and skeleton frames. In skeleton frames, the 3D positions of joints of skeleton are captured, which are also mapped into 2D silhouette images. In depth frames, each pixel contains data of player index and depth distance. It is noticed that not all the joints are tracked in some frames. We regard the frames with all joints of skeleton tracked as valid ones, and record these valid silhouette frames and their corresponding 2D and 3D skeletons. Based on these data, contours, measurements and motion data are extracted, which will be used to deform and animate virtual body models.

2.1 Contour Extraction

The front and side silhouette images of the customer are filtered from data stream of Kinect by analyzing the 3D positions of skeleton and 2D images. Then the contours of the body silhouette in front and side views are extracted by computer vision algorithm provided by OpenCV. The contours will be used to compute some measurements of the body. Since different human bodies have various detail shape, especially for the trunk of body, the contours will also be used to optimize the shape of trunk part of mannequin. There is no need to do further optimization for other part

of body because that the shape of head and limbs can be represented well by length and girth. Another reason is that the contour of limbs and head includes a lot of noisy due to the low quality of silhouette images by Kinect.

2.2 Measurement Extraction

The real stature is input by the customer, which is used as a reference to normalize the length of skeleton and other measures among valid frames into the same scale. The lengths of upper arm, forearm, thigh, and calf are then calculated by averaging the lengths of their corresponding skeleton among the valid frames. To extract the breadths of shoulder, waist, and hip, the frames with the customer captured in the front view are selected. By intersecting the contour with horizontal lines, breadths of the body silhouette are extracted. For example, the red lines in Fig. 2 are breadths of shoulder, waist, and hip of a silhouette. Although these measurements are not enough to customize template model, they can be used to predict a full set of measurements for customizing body model, which will be explained in Sect. 3.

2.3 Motion Data Collection

We use Brekel Kinect (http://www.brekel.com/?page_id=155), an application using a Microsoft Kinect and PrimeSense's OpenNI and NITE, to capture motions and stream into Autodesk's MotionBuilder in real-time, or export as BVH files. The joints of skeleton can be mapped directly onto those of the template and customized models, enabling captured motion sequences to animate models. Moreover, since there might be some noisy, it is necessary to do further smoothing to delete the frames with extraordinary or discontinuous movement if necessary.

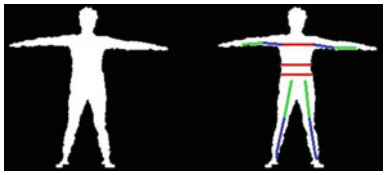


Fig. 2 A body's *front view* and some extracted measurements

3 Anthropometry-Based Body Adaptation

Anthropometric measurements are widely used as a standard in fashion industry. This section describes our anthropometry-based body adaption which customizes an animated body model for an individual person from a standard template according to partial measurements extracted from Kinect. We model the template with anthropometry in preprocessing stage and Fig. 3 shows the processes of anthropometry-based adaptation. First, a full set of anthropometric measurements is calculated from partial ones. Second, the surface and skeleton of a template are deformed according to the full set of measurements. Third, the contours of the customer are used to refine the detailed shape of the deformed body. Finally, the customized body is animated by the motion data.

3.1 Anthropometry-Equipped Template

The template is a T-pose standard human body provided by Make Human (http://www.makehuman.org). The geometry of the template model is defined by a triangular surface mesh. The skeleton and skinning information of the template is also specified.

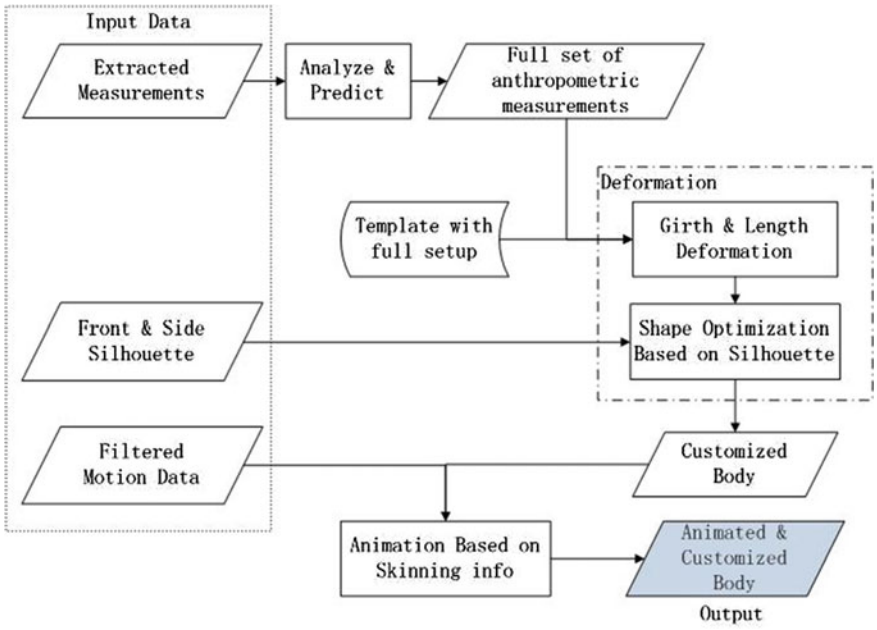


Fig. 3 Process of anthropometry based-body adaptation

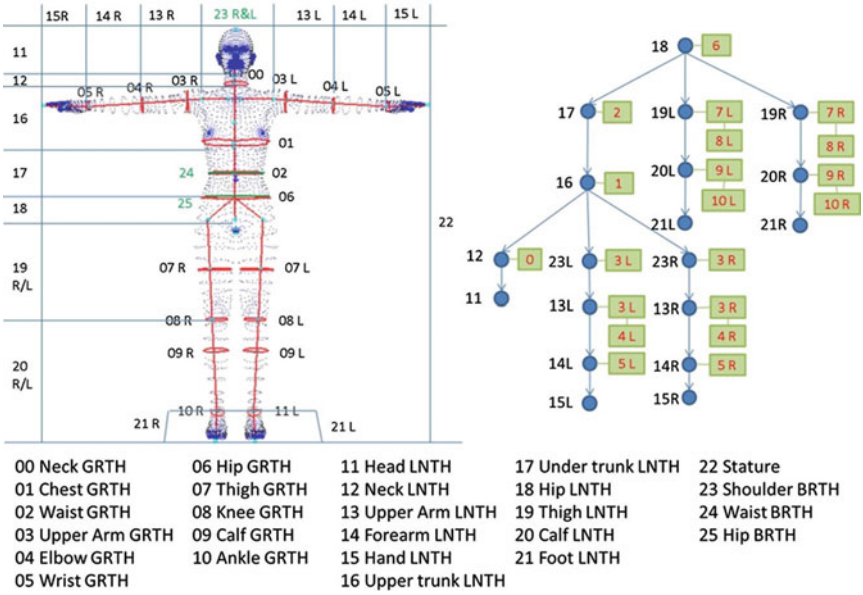


Fig. 4 Template model and structure of anthropometric measurements. *Left* Full measurement template, *Right* Tree structure of length and girth measurements, *Bottom* List of 26 measurements

While there are traditionally more than 100 anthropometric measurements for human body, we select only 26 feature anthropometric measurements as shown on the left of Fig. 4, which are commonly used in practice to characterize the body. They can be categorized into three groups:

(1) Girth G_i is measured as a circumstance of the cross-section in the plane perpendicular to the skeleton of a local part of body.
(2) Length L_i is specified as a distance between two particular points or a distance between two cutting planes.
(3) Breadth B_i is measured as the horizontal breadth of a body segment at the level of specific joint point.

It can be seen that these girth, length and breadth measurements are nicely linked with the template mesh, joints and skeleton. Thus a full setup body template consists of the mesh, 26 anthropometric measurements, skeleton and skinning information.

3.2 Full Measurement Calculation

From Kinect, we can obtain partial measurements based on which we will estimate the full set of measurements. In the following we present a prediction method to achieve this.

To demonstrate our idea and approach, in this chapter we use ANSUR anthropometric body measurement survey (ANSUR) as our training data set, which contains anthropometric measurements for 2,208 females. For each female in the training set, the 26 feature measurements are extracted and they form a feature vector to represent the female. Let $a_i=[a_{i,1}, ...,a_{i,j}, ...,a_{i,26}]$ be the feature vector of female i, where $a_{i,j}$ is its j-th measurement. Then, the training data for anthropometry analysis can be represented by $A = [a_1, a_2, ...,a_n]^T$, in which n is the count of the training set. With reference to Fig. 4.4, each feature vector a_i should satisfy the constraint C: $a_{i,11} + a_{i,12} + a_{i,16} + a_{i,17} + a_{i,18} + a_{i,19} + a_{i,20} = a_{i,22}$. Constraint C means that the height of the body equals the sum of lengths of head, neck, upper trunk, under trunk, thigh and calf.

Using the approach described in Sect. 2, we can extract some measurements from Kinect. We let $w_{(1)}=\{w_{13}, w_{14}, w_{19}, w_{20}, w_{22}, w_{23}, w_{24}, w_{25}\}$ denotes these known measurements for a customer and $w_{(2)}$ denotes the rest unknown measurements. Thus our task here is to calculate $w_{(2)}$ with constraint C.

The correlation coefficient between measurements i and j, Cor_{ij}, which measures the influence of one anthropometric measurement to another, can be calculated by:

$$Cor_{ij} = \frac{\sum_{k=1}^{n} \left(a_{k,i} - \overline{a_i}\right)\left(a_{k,j} - \overline{a_j}\right)}{\sqrt{\sum_{k=1}^{n} \left(a_{k,i} - \overline{a_i}\right)^2} \sqrt{\sum_{k=1}^{n} \left(a_{k,j} - \overline{a_j}\right)^2}} \tag{1}$$

where $\overline{a_i}$ and $\overline{a_j}$ are the mean values of measurements i and j, respectively. The linear correlation between $w_{(1)}$ and $w_{(2)}$ is modeled by a matrix $Cor = [Cor_{ij}]$. *Singular value decomposition (SVD)* is performed to obtain

$$Cor = U \sum V^T \tag{2}$$

where U and V are orthonormal bases for $w_{(1)}$ and $w_{(2)}$, and $\sum$ reflects the correlative relation between the two bases. The calculation of $w_{(2)}$ is done by the following steps:

(1) Assume $w_{tpl(1)}$ and $w_{tpl(2)}$ are the sub feature vectors of the template corresponding to $w_{(1)}$ and $w_{(2)}$, respectively. Then $w_{tpl(1)}$ is projected into the space with bases of U to get $u = U^{-1}w_{tpl(1)}$. Similarly, $w_{(1)}$ is projected into the space with bases of U to get $u' = U^{-1}w_{(1)}$. Based on this, we could detect the variation in U space, which is $\Delta u = u - u'$

(2) We project $w_{tpl(2)}$ into the space with bases of V to get $v = V^{-1}w_{tpl(2)}$. Since $\sum$ reflects the correlative relation between the two bases, the variation of $w_{(2)}$ in

V space, $\Delta v_k = \Delta u_k$ if $\sum_k^T > 0.5$ (a threshold we choose); and $\Delta v_k = 0$ otherwise. Then the value of $w_{(2)}$ in V space can be obtained by $v' = v + \Delta v_k$.

(3) $w_{(2)}$ can be calculated by projecting back from V space, $w_{(2)}$ = Vv'.

(4) So far $w_{(2)}$ is predicted based on $w_{(1)}$ through correlative analysis. A post processing step is needed to further modify the measurements to satisfy constraint C.

Let $\Delta_{22} = w_{22} - (w_{11} + w_{12} + w_{16} + w_{17} + w_{18} + w_{19} + w_{20})$ and S={11, 12, 16, 18, 19, 20}. The value of measurements w_i, whose index is in S, is updated by

$$\frac{w_i}{\sum_{k \in S} w_k} \Delta_{22} + w_i \tag{3}$$

Now a full set of anthropometric measurements w which satisfy constraint C is obtained.

3.3 Anthropometry-Based Deformation

We now describe how to deform the template mesh model to match a given set of anthropometric measurements.

We first organize the measurements in a proper structure. For each girth measurement G, a set of intersection points P_G and edges E_G for the cross-section can be obtained by intersecting the body mesh with the cutting plane. G is represented by the circumference of a sequential intersecting point list of P_G. P_G is considered as the feature constraint of G. For length measurements which correspond to skeleton, a tree structure is organized based on the skeleton. Each length measurement is associated with an inner skeleton and one or several related girth measurements (see the right of Fig. 4 for an illustration).

Then we deform feature vertices through length and girth deformation. The length deformation is to linearly scale the corresponding inner skeleton and to relocate the related girth measurements. Based on the tree structure, the deformation of one node will cause translation of its descendants. For girth deformation, if the value of girth G is changed from g to g', then each feature point P_i in P_G is moved to a new location $p_i' = c + \frac{g'}{g}(p_i - c)$ to achieve the target value of girth, where point c is the center of P_G.

Next, the vertices of the template mesh are deformed through the Laplacian reconstruction (Lipman et al. 2004; Nealen et al. 2005) with some constraints at feature points P_G. For each vertex v_i, its Laplacian coordinates is defined by:

$$\delta_i = v_i - \sum_{j \in N(i)} w_{ij} v_j \tag{4}$$

where $N(i)$ is the set of indices of the vertices that share an edge with v_i, and w_{ij} is determined by cotangent weights (Meyer et al. 2002) satisfying $\sum_{(i,j)\in E} w_{ij} = 1$. Let v be the column vector consisting of all vertices of the mesh and *delta* be the column vector consisting of the Laplacian coordinates of all the vertices. Then $Lv = \delta$, where $L = I - A$ with $A_{ij} = \begin{cases} w_{ij} & i \in N(j) \\ 0 & otherwise \end{cases}$ Assume initially the feature point P_k in P_G is on the edge between vertices v_i and v_j. That is, $P_k = (1 - \lambda_k) * v_i + \lambda_k * v_j$ with constant λ_k. After the girth and length deformations, all the feature points P_k in P_G have been updated into P'_k. Then the reconstruction problem is to find new vertex vector v' for v of the mesh, which minimizes the objective function:

$$E(V') = \| \delta - L(v') \|^2 \tag{5}$$

Subject to $(1 - \lambda_k) * v_i + \lambda_k * v_j = P'_k$ for each P_k in P_G. This objective function aims to make the change of the Laplacian of each vertex of the mesh before and after the deformation as small as possible. Since the Laplacians reflect the geometric shape well, the Laplacian reconstruction usually produce nice deformation.

3.4 Shape Refinement

To use the front and side contours to optimize the shape of waist part of deformed model, the corresponding contours on 3D model and how to optimize should be detected. First, the contours are normalized into the same scale with 3D model. Second, four contour lines of model are detected by intersecting model horizontally and vertically as showed in 3rd column of Fig. 5.

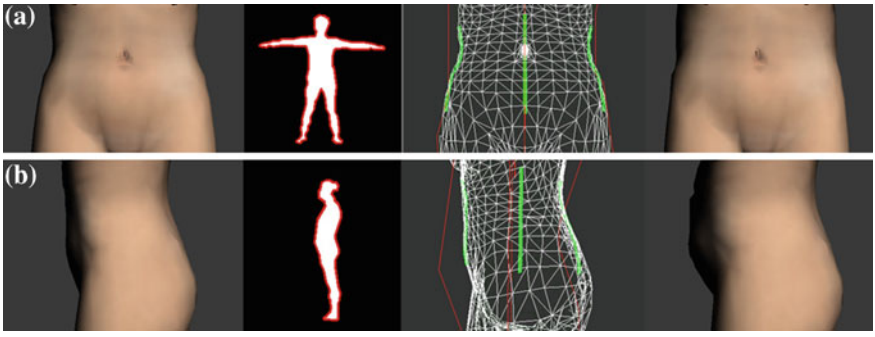


Fig. 5 Shape refinement. From *left* to *right* Initial shapes, 2D contours, 3D contours, refined shapes

(1) Left and right waist contours: intersect model by x–y plane with z equals z value of point at widest of waist girth and extract line below under breast plane and above hip plane.
(2) Front and back waist contours: intersect model by y–z plane with x equals middle of body model and extract line below under breast plane and above hip plane.

Then, each point P_i in these four contours is regards as a constraint $C\_contour_i$. For points on left and right waist contours, the x-offset of point can be calculated by $1/2\Delta width$, in which $\Delta width$ equals the difference between width of model and width of 2D front contour. For points on front and back waist contours, we first project corresponding skeleton into 2D side contour. Then the z-offset of each point is $\Delta thickness$ of difference between thickness of model and target 2D contour, in which thickness means distance from point to skeleton.

The shape refinement is performed by applying the Laplacian reconstruction again on the vertices in related region including upper trunk, under trunk, and thigh parts, with the constraints imposed by the points on the contours. The fourth column of Fig. 5 is the result of refining the initial shape (the first column) according to the contours in the second column.

3.5 Animation of the Customized Model

Since the skinning information is attached in template and the embedded skeleton has been deformed during the length deformation, the standard linear blend skinning (LBS) method is thus used to animate the customized body model. Based on LBS, the position of the transformed vertex j is calculated by

$$v'_j = \sum_i w_j^i T^i(v_j) \tag{6}$$

Where T_i is the transformation of the i-th bone, and w_{ij} is the weight of the i-th bone for vertex j. Figure 6 compares the real frames of a customer and her animated virtual mannequin guided by Kinect motion data.

4 Garment Prepositioning and Simulation

Once the customized body is generated, the next task for virtual try-on is to dress the body with clothes that are chosen from garment database. This usually involves two steps. The first one is prepositioning, which is to map 2D patterns of garment onto the 3D body and generate initial clothes. The second one is simulation, by which the initial clothes are deformed dynamically to fit the animated body.

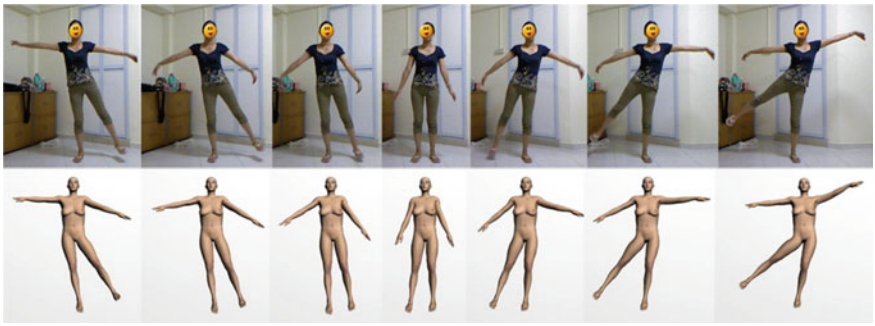


Fig. 6 A real customer and her corresponding animated model

Prepositioning is an important step. A good prepositioning result will reduce the computation of simulation; otherwise, a bad prepositioning will require high computation cost or even leads to failure of simulation as shown in Fig. 7.

To the best of our knowledge, so far the prepositioning is conducted manually or semi-automatically. This is because based on the geometric information of patterns only, it is in general impossible to preposition the patterns around body automatically without interaction. Here we propose an approach to achieve automatic prepositioning. Our basic idea involves four aspects: first, we introduce a new 2D pattern format for garment, which includes geometry, connectivity and semantic information; second, as shown in Fig. 8, the body template is used as a reference to link the correspondence between 2D patterns and the customized body model; third, the 2D patterns are deformed and refined according to the correspondences of skeleton and constraints of seams; and fourth, the 2D patterns are merged into an initial 3D garment by topological stitching along the seams.

4.1 Data Format

Currently garment is defined by 2D patterns in various industry formats. We propose to embed connectivity and semantics knowledge into the existing geometry-based format in order to make automatic prepositioning possible. In our

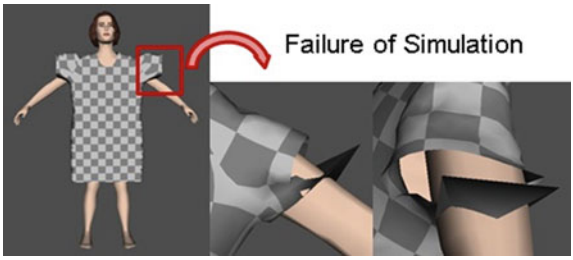


Fig. 7 Example of bad prepositioning which leads to failure of simulation

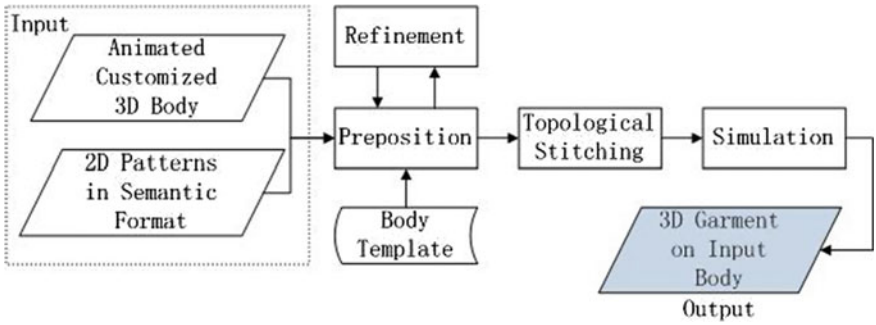


Fig. 8 The process of garment preposition and simulation

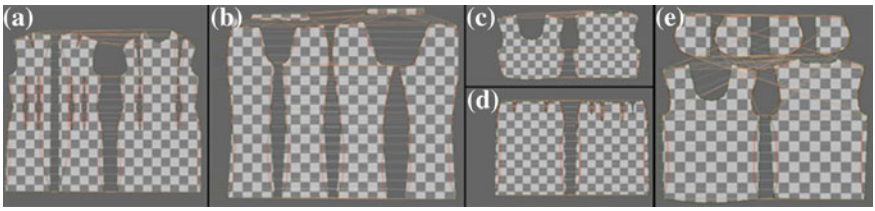


Fig. 9 2D Patterns of garments in category

proposed format, three types of information are involved for 2D patterns of a garment:

(1) Geometrical information: 2D patterns (2D pieces in Fig. 9) are defined by 2D contours and triangulated into 2D mesh.
(2) Topological information: Seams (red lines in Fig. 9) specify the linkage along the contours of one or two patterns.
(3) Semantic information: There are some tags for layer id, side id, etc. The markers to specify with which skeleton of the template the patterns are linked and initial 3D position placed on the template for 2D patterns are recorded.

While geometry is used to define the shape of individual patterns, topological and semantic information is embedded and used to indicate the relationship among patterns. Before prepositioning, there is a stage of preprocessing done previous offline. The goal of preprocess is to define and record semantic information. The patterns of garment are positioned onto the template model by designers either by manually or other tools. We also design a tool for automatic prepositioning of basic types of garment. For complicated garment, it need designers to place or modify manually. After this stage, each pattern P corresponds to a skeleton S_{id} of the template model and a transformation matrix T_{tpl} which places the pattern onto the template.

4.2 Prepositioning

First, since the customer's model is adapted from template, the patterns on the template can be deformed accordingly. Assume that the matrix for deforming the skeleton of the template to that of the customer is T_{skl}. This transformation is applied to the patterns.

Second, by checking the bounding volume for each body segment of the skeleton and considering the layer id of patterns, a translation matrix T_{vol} is applied on patterns to avoid intersection with body and other patterns.

Third, there are three kinds of intersection that should be eliminated through iterations. The first is the intersection of seam lines with the body, which can be detected by testing seams. The other two are the intersection between the pattern and the body and the intersection between two patterns, which can be detected by testing the mesh of patterns and the body model. To eliminate these intersections, the involved patterns are transformed along the direction of the skeleton by analyzing the information of connectivity and geometry. We denote the required transformation by T_{ref}.

To combine these steps, the new position of a pattern on customer's model, P', can be computed from the initial position P by

$$P' = T_{\text{ref}} \times T_{\text{vol}} \times T_{\text{skl}} \times T_{\text{tpl}}(P) \tag{7}$$

The second image in Fig. 10 is the positioned result from initial placement on the template shown in the first image.

4.3 Topological Stitching and Simulation

Assembly of patterns is to stitch the meshes of patterns together along the seams. In the process of stitching, the most important thing is to refine and calculate right topology of garment. First, we push the mesh of all patterns, which includes vertices and faces information, into a temp vertex and a temp face list of garment. Second, since each seam corresponds to two lists of related vertices, we choose one of the two as an updated list, and the other as a reference. For each vertex in the updated list, we detect faces which contain the vertex and modify the vertex index of faces as the index of the reference list. Then we mark the vertices in the updated list as invalid. Third, the valid vertices in the temp vertex list are copied into the vertex list of garment. The vertex index of faces in the temp face list is updated and copied into the face list. The right of Fig. 10 shows the result of stitching. Although the initial 3D garment has a flat shape of garment, it includes not only the geometrical information but also the topological information. In this way, the prepositioning is simply accomplished, without complicated user interaction.

Finally, the simulation of the clothes on the animated body is performed. There have been many research works done on this and some libraries have been

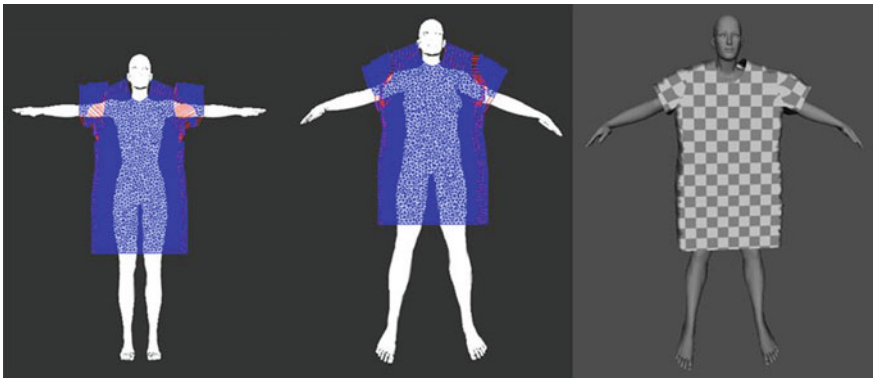


Fig. 10 Example of prepositioning and topological stitching. *left* patterns on the template model; middle: prepositioning on the customized model, *right* topological stitching result

developed to simulate clothes when they are initially placed around the body (Magnenat-Thalmann and Volino 2005; Magnenat-Thalmann 2010). We use the library in Fashionizer (Magnenat-Thalmann 2010), which is a virtual garment creation and simulation software developed by MIRALab, to do the simulation to generate the final dressed 3D body.

5 Experimental Results

Figure 11 is an illustration of our user interface. The customer should show their valid front and side view first. Then, they are required to move and pose in front of Kinect to do motion capture. After that, a virtual body model for the customer will be generated and displayed. For virtual try-on, the customer could pick up one or more garments from the category in the right. A simulation on static model will be shown quickly first, and the final try-on result with motion will be displayed after simulating garment on whole frames.

5.1 Evaluation of Body Generation

To evaluate body generation, we invite several female volunteers of different ages, shapes, and statures to try this system. Experimental results of three customers are compared and listed in Fig. 12.

We compare the three customized models and their real photos in the front and side view, respectively. The stature of customer A and B are nearly the same and A is fatter than B as showing in real photo. Customer C is taller than A and B. Our customized models also reflect this observation. The abdomen of customer C is bulge in photo and the virtual model also owns the same shape with the help of shape optimization.

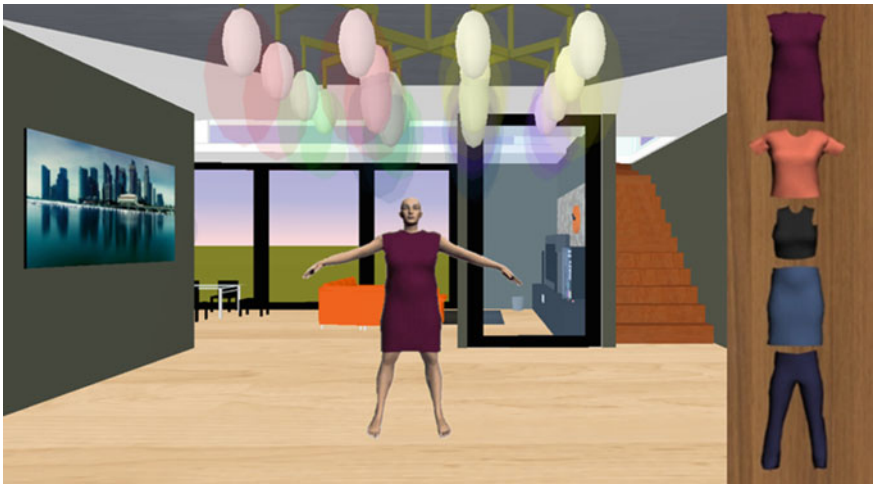


Fig. 11 User interface of our framework

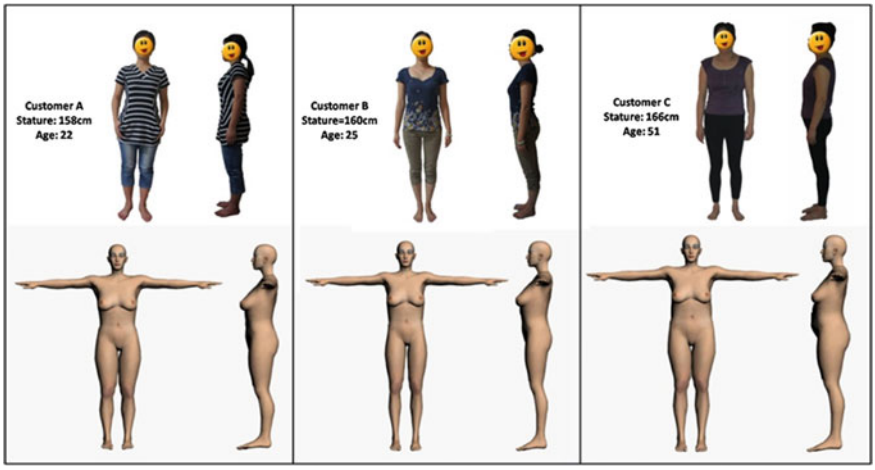


Fig. 12 Experimental results of body generation. 1*st row* photos and information of customers, 2*nd row* virtual models for customers in the front and side views

5.2 Experiment of Garment Try-on

To evaluate virtual try-on effect, we preposition and simulate different set of garments on customers. Several garments of different types are chosen, which include a pair of trousers, a skirt, a T-shirt, a mini-top, and dresses. As shown in Fig. 9, we embed the seams and connectivity into the 2D patterns and record them in our format previously.

Fig. 13 Experimental results of garments prepositioning: initial positioned and stitched garments on models

The experimental results of prepositioning are illustrated in Fig. 13. All the five basic types of garments can be placed on different body model in a good way, since the simulation of initial 3D garments are converged with no failure.

To evaluate the final try-on effect, the experimental results of movement sequences are illustrated in Fig. 14. Customer A dressed with garment (e) and (d) are turning right, and Customer B with garment (c) and (b) walks straight forward, while customer C dressing on garment (a) are turning around. The experimental results demonstrate that our try-on system with Kinect can approximate customer

Fig. 14 Experimental virtual try-on effect of simulated garments on animated models **a** Customer A; Garment e and d **b**Customer B; Garment c and b **c**Customer B; Garment c and b

by generating virtual body in a convenient way, provide accurate preposition results and precise simulation of try-on garments.

Compared with (Cordier et al. 2003), our system simplifies the process of body generation, with which the customer need not to input all the measurements. The information captured by Kinect is enough to generate an accurate virtual model. When comparing our system with Bodymetrics, a sequence of frames with garment dressing on animated model could be displayed in our system, instead of only a static simulation result.

6 Conclusion

This chapter has described a Kinect-based framework for virtual try-on, which includes 3D body model generation and cloth dressing. The framework uses an anthropometry-based modeling template to build a bridge to link the customized model and the patterns of garment. For the body generation part, the main contribution lies in a feasible statistical method to predict a full set of measurements and a robust method through anthropometry-based deformation and contour-based shape refinement, which can produce smooth body models efficiently. By analyzing the data extracted from Kinect, the partial measurements and motion data can be used to make the whole virtual try-on process simple. For the garment dressing part, the main contribution includes defining the format for connectivity and semantic information and proposing a prepositioning algorithm.

Using the proposed framework, a customer should wear tight clothes so that accurate contours and skeleton can be extracted. Since anthropometry-based analysis predicts "normal" results of measurements, some customers with very irregular characteristics need to manually edit the measurements. Since garment simulation is computationally intensive, the whole system cannot achieve real time right now. In addition, some other aspects such as skin and face adaptation and garment resizing are also very interesting topics for future work.

Acknowledgements The authors would like to acknowledge the Ph.D. grant from the Institute for Media Innovation, Nanyang Technological University, Singapore.

References

Allen B, Curless B, Popovi′c Z (2003) The space of human body shapes: reconstruction and parameterization from range scans. ACM Trans Graph 22(3):587–594

ANSUR. http://www.dtic.mil/dticasd/docs-a/anthro_military.html

Barron C, Kakadiaris IA (2000) Estimating anthropometry and pose from a single image. IEEE Comput Soc 1:669–676

Bodymetrics. http://www.bodymetrics.com/

Brekel Kinect. http://www.brekel.com/?page_id=155

Cordier F, Seo H, Magnenat-Thalmann N (2003) Made-to-measure technologies for an online clothing store. IEEE Comput Graph Appl 23:38–48

Fontana M, Rizzi C, Cugini U (2006) A cad-oriented cloth simulation system with stable and efficient ode solution. Comput Graph 30:391–406
Fontana M, Carubelli A, Rizzi C, Cugini U (2008) Clothassembler: a cad module for feature-based garment pattern assembly. Comput Aided Des Appl 2(6):795–804
Fuhrmann A, Groß C, Luckas V, Weber A (2003) Interaction-free dressing of virtual humans. Comput Graph-Uk 27(1):71–82
Groß C, Fuhrmann A, Luckas V (2003) Automatic pre-positioning of virtual clothing. In: Proceedings of the 19th spring conference on Computer graphics, SCCG'03, pp 99–108
Grosso MR, Quach RD, Otani E, Zhao J, Wei S, Ho PH, Lu J, Badler NI (1987) Anthropometry for computer graphics human figures, Technical Report No. MS-CIS-89-71
Izadi S, Newcombe RA, Kim D, Hilliges O, Molyneaux D, Hodges S, Kohli P, Shotton J, Davison AJ, Fitzgibbon A (2011) KinectFusion: real-time dynamic 3D surface reconstruction and interaction, ACM, New York, p 23
Jianhua S, Magnenat-Thalmann N, Thalmann D (1994) Human skin deformation from cross-sections. In: Proceedings of the computer graphics international, pp 9–24
Kasap M, Magnenat-Thalmann N (2010) Customizing and populating animated digital mannequins for real-time application. In: Proceedings of the 2010 international conference on cyberworlds, CW'10, pp 368–374
Kasap M, Magnenat-Thalmann N (2011) Skeleton-aware size variations in digital mannequins. Vis Comput 27:263–274
Lipman Y, Sorkine O, Cohen-Or D, Levin D, Rössl C, Seidel HP (2004) Differential coordinates for interactive mesh editing. In: Giannini F, Pasko A (eds) Shape modeling international 2004 (SMI 2004), pp 181–190
Magnenat-Thalmann N (2010) Modeling and simulating bodies and garments. ISBN 978-1-84996-263-6, Springer-Verlag Berlin, Heidelberg
Magnenat-Thalmann N, Volino P (2005) From early draping to haute couture models: 20 years of research. Vis Comput 21(8–10):506–519
Magnenat-Thalmann N, Seo H, Cordier F (2004) Automatic modeling of virtual humans and body clothing. J Comput Sci Technol 19(5):575–584
Maïm J, Yersin B, Thalmann D (2009) Unique character instances for crowds. IEEE Comput Graph Appl 29:82–90
Make Human. http://www.makehuman.org
Meyer M, Desbrun M, Schröder P, Barr AH (2002) Discrete differential-geometry operators for triangulated 2-manifolds. In: Hege HC, Polthier K (eds) Visualization and mathematics III, Springer, Berlin, pp 35–57
Nealen A, Sorkine O, Alexa M, Cohen-Or D (2005) A sketchbased interface for detail-preserving mesh editing. ACM Trans Graph 24:1142–1147
Seo H, Magnenat-Thalmann N (2004) An example-based approach to human body manipulation. Graph Models 66(1):1–23
Seo H, Yeo YI, Wohn K (2006) 3d body reconstruction from photos based on range scan. In: Proceedings of the first international conference on technologies for e-learning and digital entertainment, Edutainment'06 pp 849–860
Thanh TL, Gagalowicz A (2005) Virtual garment pre-positioning. In: Proceedings of the 11th international conference on computer analysis of images and patterns, CAIP'05, pp 837–845
Thanh TL, Gagalowicz A (2009) From interactive positioning to automatic try-on of garments. In: Proceedings of the 4th International conference on computer vision/computer graphics collaboration techniques, MIRAGE'09, pp 182–194
Volino P, Magnenat-Thalmann N (2000) Virtual clothing-theory and practice. ISBN 978-3-642-57278-4, Springer, London

Integrating EEG Modality in Serious Games for Rehabilitation of Mental Patients

Chun Siong Lee, Chee Kong Chui, Cuntai Guan, Pui Wai Eu, Bhing Leet Tan and Joseph Jern-Yi Leong

Abstract In order to develop an effective autonomous rehabilitative training solution with serious games for use with occupational therapy of mental patients, the authors postulate that a solution could come from combining human performance engineering with cognitive science. Conventional physical task performance metrics are insufficient in assessing the overall performance of the mental patients. Ideally, the mental state of the subject should be assessed in conjunction with the motor performance in order to better evaluate the rehabilitation progress of mental patients. In order to observe and evaluate the mental state of the subjects, noninvasive scalp EEG was used. The EEG readings were then analyzed using Lempel–Ziv complexity (LZC) and spectral analysis. The objective of this study is to compare task performance against the corresponding EEG analysis that enables the identification of possible neural markers for gauging mental activity pertinent to the mastery of simple tasks. Results identify that LZC values and activity in the theta and low alpha frequency band of the spectral analysis in the central, occipital, and parietal regions can possibly be used as the neural markers.

Keywords EEG · Serious games · Rehabilitation · Mental patients

C. S. Lee (✉) · C. K. Chui
National University of Singapore, Singapore, Singapore
e-mail: csl@nus.edu.sg

C. Guan
Institute for Infocomm Research, Singapore, Singapore

P. W. Eu · B. L. Tan · J. J.-Y. Leong
Institute of Mental Health, Singapore, Singapore

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_4,

1 Introduction

In a large-scale epidemiological study conducted in 2010 by the Institution of Mental Health in Singapore, it was found that more than 1 in 10 people will be stricken by a form of mental illness within their lifetime. Increasing prevalence of mental illness in society is a growing concern and as a result, there is a need for better management and rehabilitation of these mental patients. In order to reintegrate mental patients back to society as productive and self-reliant individuals, mental patients undergo rehabilitation and are exposed to occupational therapy that are designed to train them for suitable jobs in areas such as clerical work and food handling in food and beverage (F&B) industries. Conventionally, due to the delicate nature of the interactions with mental patients, occupational therapists are required to manually train them in cognitive and motor skills crucial in the job scope. In addition, it is often hard to judge a subject's level of competence and task mastery without the professional supervision of therapists. As such, the occupational therapy sessions are limited to small groups. These factors put pressure on the already constrained manpower available. Therefore, one possible solution is to create an autonomous rehabilitative training protocol with serious games that mental patients can train independently so as to reduce the reliance on therapists.

Conventionally, serious games involves simulations of tasks with the main notion of educating the user to master achieving certain task objectives, likely through a supervised or reinforcement learning protocol. For example, the occupational task objectives for a kitchen aide could be the proper technique for wielding and using a knife. The corresponding task performance would be the speed and accuracy of cutting the food ingredients. However, with respect to the training of mental patients, purely using conventional task performance metrics are insufficient in assessing the overall performance of the mental patients. Revisiting the example of a kitchen aide's cutting skill, an occupational therapist would focus more on the confidence in wielding the knife and mood of the mental patient in participating in the training before the product of the practice itself. Hence, the focus is on affective (Li et al. 2009) nurturing of the mental state of the patient that further enables effective learning of tasks. This example highlights that in order to develop an effective rehabilitation solution, we should combine human performance engineering with cognitive science. Hence, to assess the rehabilitation progress of mental patients, the motor performance and mental state should ideally be assessed in conjunction with each other.

The most common way of observing a person's mental state would be through noninvasive scalp EEG readings. However, EEG itself is an indirect measure of mental activity through the spatial recording of spontaneous electrical fields that perpetuate a person's brain which are generated by the ionic activations of neurons in the brain. Raw EEG data can be recorded with multiple low impedance electrodes on the scalp. However, in order to interpret the sheer multiplicity of EEG readings, post processing of the EEG data is required.

A conventional way of post processing the EEG data is through the Fourier transformation of the signals into the frequency domain. This is because neurons are known to activate with causal synchrony resulting in neural oscillations across the brain. This synchronous activity is revealed with frequency analysis of the EEG data. The classification of brain waves into frequency bands within the frequency domain has been classically shown to correlate with certain mental states and spatial prevalence (Friel 2007). For example, the alpha frequency band in the occipital region (8–13 Hz) has been associated with a more relaxed mental state (Klimesch et al. 1993; Aftanas et al. 1996). Hence, the relative changes in activity in the frequency domain are an indication of the mental state of the brain as the subject performs the training task. In fact, one of the earliest clinical uses for EEG was in EEG biofeedback, which was fundamentally based on training individuals to alter levels of brain activity in certain frequency bands. This form of therapy was widely used with the treatment of attention deficit/hyperactivity disorder (ADHD) where ADHD patients were frequently found to have relatively elevated levels of activity in the theta frequency band (4–7 Hz) and lower relative levels of activity in the alpha (8–13 Hz) and beta (14–25 Hz) range (Monastra et al. 2005).

Another method of interpreting the raw EEG recordings is through the computation of the complexity of the EEG time series data. Lempel–Ziv Complexity (Lempel and Ziv 1976) has been a popular method of characterizing several types of biological signals including classifying genomic sequences (Gusev and Nemytikova 1999; Chen et al. 2000), identifying subjects with Alzheimer's disease (Abásolo et al. 2006), classifying fatigue (Zhang and Zheng 2009) and the depth of anesthesia from EEG complexity (Zhang et al. 2001). It represents the rate of emergence of new patterns in a time series. Hence, a higher LZC value indicates that the signal is more complex. Before an EEG signal can be computed with LZC, it first has to be converted into a binary signal. Conventionally, the median or mean of the signal is used as the binary thresholding value (Nagarajan 2002). Data points above the thresholding value are converted into ones and data points below that value are converted into zeroes. The binary string is then read from left to right while each character is compared against a "vocabulary of words." As the string is read, the vocabulary increases and the rate of unique word entry into the vocabulary represents the emergence of new patterns in the string. Normalized over a function based on the length of the entire string, this value indicates the complexity of the string. Thus, LZC is useful in reducing the dimensionality of an EEG data channel into a single value.

In this study, the objective is to compare a subject's task performance against the corresponding scalp EEG data recorded during the task. The comparison enables the identification of possible markers for gauging mental activity pertinent to the mastery of simple tasks such as the origami box folding used in this study. This enables a benchmark for deeper analysis of a person's performance beyond that of the conventional physical kinematic analysis.

2 Methodology

2.1 Subjects and Experimental Protocol

We recruited six healthy male volunteers with ages between 21 and 30 who were all right handed to perform the study. The experimental task chosen in this study was to fold a simple paper origami box. The folding instructions were presented in the form of 39 animated steps within a PowerPoint presentation. Each subject was required to build one box per trial with five trials in all. The first four trials were aided by the PowerPoint instructions while the last trial was unassisted.

2.2 Equipment

In order to record the scalp EEG readings, a Neuroscan NuAmps EEG amplifier was used. The amplifier is connected to a flexible lycra stretch cap where 32 sintered silver/silver chloride electrodes have been pre-positioned to the International 10–20 system of EEG electrode placements. The amplifier converts the analog electrode signal and outputs to a PC for recording. A total of 19 electrode channels were recorded in this study, and their topographical map is shown in Fig. 1. EEG conductive gel were applied on each electrode to reduce the impedance of the electrodes (<5 KΩ) before the EEG data was recorded. On a separate PC, a series of animated origami folding instructions are shown to the subject (see Fig. 2). A webcam was also used to record the subjects performing the task (see Fig. 3).

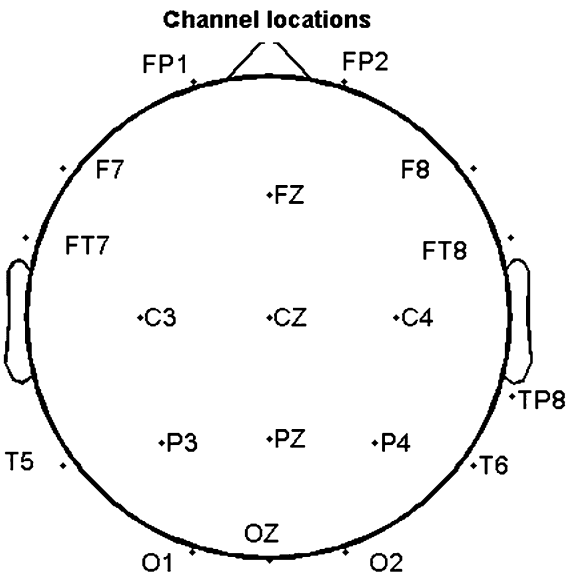


Fig. 1 Topographic plot of the EEG electrodes recorded in accordance to the international 10–20 system of EEG electrode placement and labeling

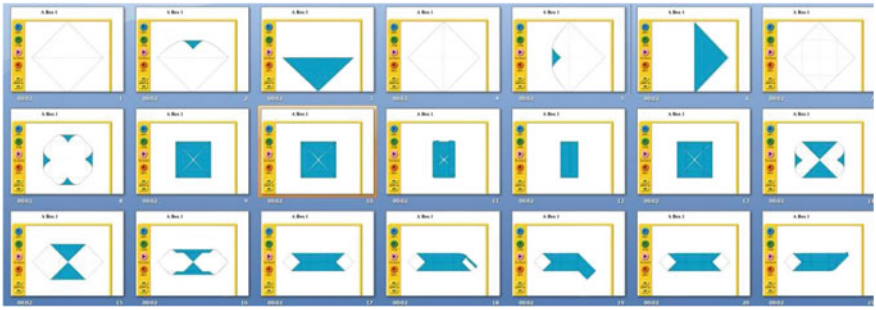


Fig. 2 Sequential screenshots of the origami folding instructions shown to the subjects

Fig. 3 Screenshot of subject 1 performing the folding task

2.3 EEG Processing

In this study, raw EEG data was recorded from the 19 EEG channels sampled at 1,000 Hz and exported for further processing in the well-known EEGLAB toolbox (Delorme and Makeig 2004) for Matlab. Using EEGLAB, the EEG data was notch filtered at 50 Hz to remove electrical line noise. Recorded EEG data tends to be very noisy due to external physiological artifacts such as ocular motion (EOG) and muscle activation (EMG). Eye blinks also induce large slow waveforms at the frontal regions of the EEG recorded. These obvious visible noise artifacts are manually cropped from the raw EEG readings before analysis. The EEG signal is then bandpass filtered between 0.5 Hz and 45 Hz to remove low frequency noise such as slow drift found in continuous raw EEG data and high frequency noise such as muscle activations.

The Fast Fourier transform (FFT) function implemented in Matlab was used by EEGLAB to perform the spectral analysis of the EEG data. Subsequently, for LZC calculation, the EEG data after preprocessing is converted to a binary string by

thresholding using the median value of the individual channel's signal. The binary string is then fed into a Matlab function to extract the LZC value.

3 Results

The six subjects took an average of 14 min to complete the five trials. The individual time taken for all six subjects to perform the five folding tasks is plotted in Fig. 4. It can be seen that the subjects were able to almost plateau the time taken to fold the paper box by the third trial. The timings for trial 5 remained low, suggesting that the subjects have managed to master the folding task by the fourth iteration since there was no guidance given in trial 5. Looking at Fig. 4, it is evident that subject 5 was the worst performing subject overall and subject 6 was the fastest performer overall. Even with the relatively stabilized plateau of timings between trials 3–5, subject 5 still took approximately three times as long as the fastest person, subject 6, to perform the folding task in those trials. In the subsequent spectral analysis, we shall focus mainly on the disparity between the trends in these two subjects' EEG data.

The calculated LZC values for each subject are plot for the 19 EEG channels in Figs. 5, 6, 7, 8, 9 and 10. The LZC values averaged across the six subjects is shown in Fig. 11. The LZC values are found to range between 0.10 and 0.35 for all the EEG data recorded in this study.

The overall topological frequency spectrum plots of subject 5 over trial 1–5 are shown in Fig. 12a–e. Similarly, the topological frequency spectrum plots for subject 6 are shown in Fig. 13a–e. The individual frequency spectrums of channels C3, C4, CZ, O1, OZ, O2, P3, PZ, and P4 are plotted in Fig. 14a–i and Fig. 15a–i for subject 5 and subject 6, respectively.

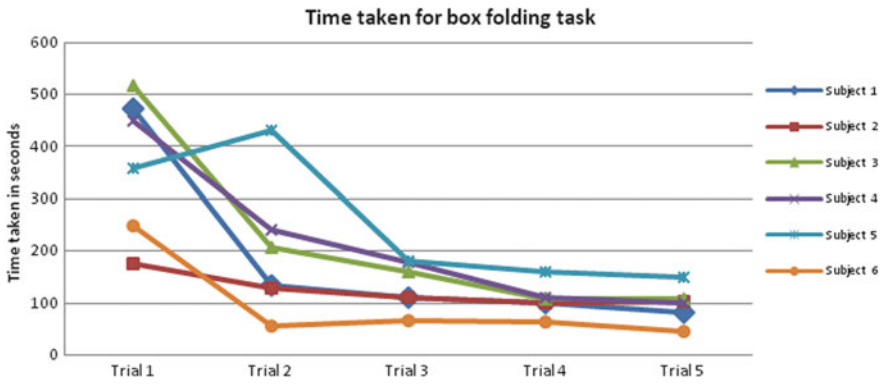


Fig. 4 Time taken for each origami box folding trial of all six subjects

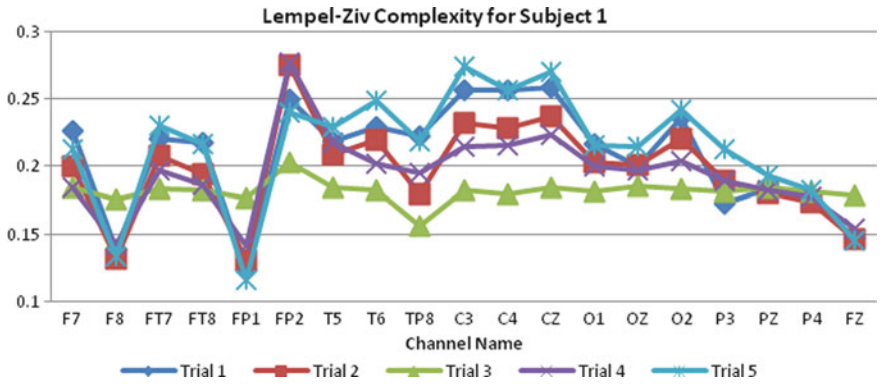


Fig. 5 Lempel-Ziv complexity distribution for all 19 EEG channels and all five trials of subject 1

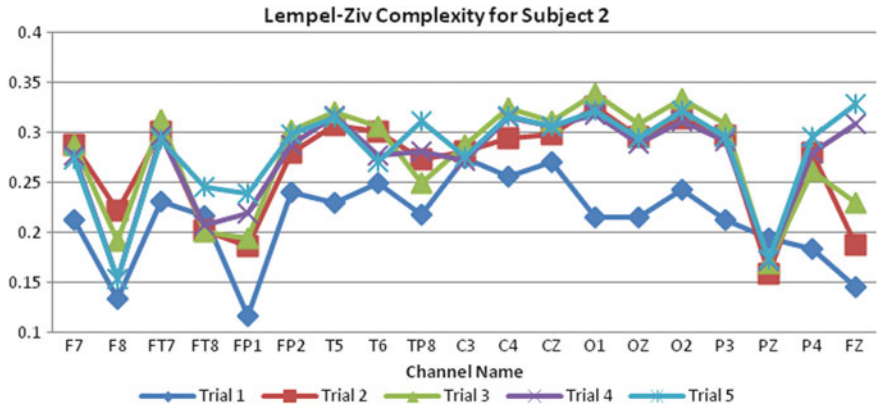


Fig. 6 Lempel-Ziv complexity distribution for all 19 EEG channels and all five trials of subject 2

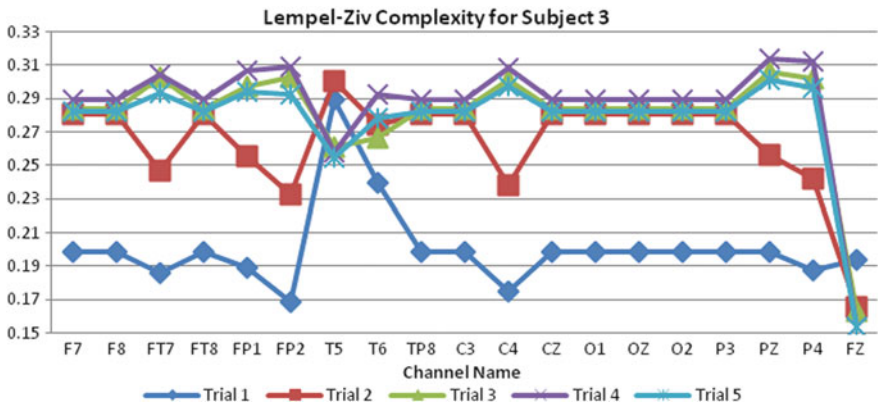


Fig. 7 Lempel-Ziv complexity distribution for all 19 EEG channels and all five trials of subject 3

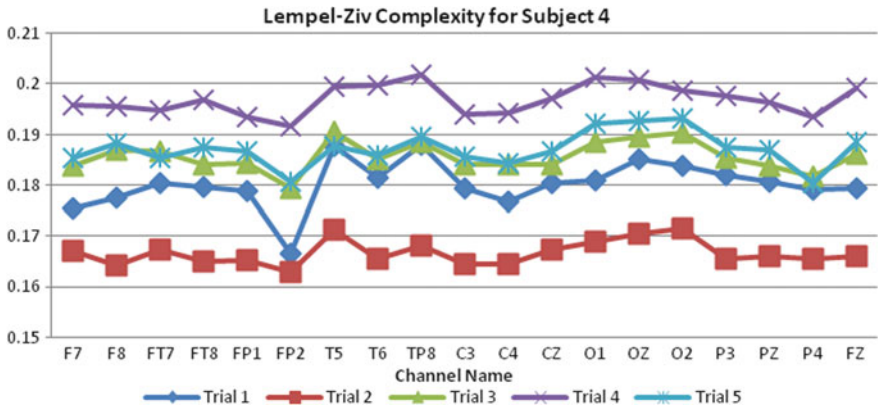


Fig. 8 Lempel-Ziv complexity distribution for all 19 EEG channels and all five trials of subject 4

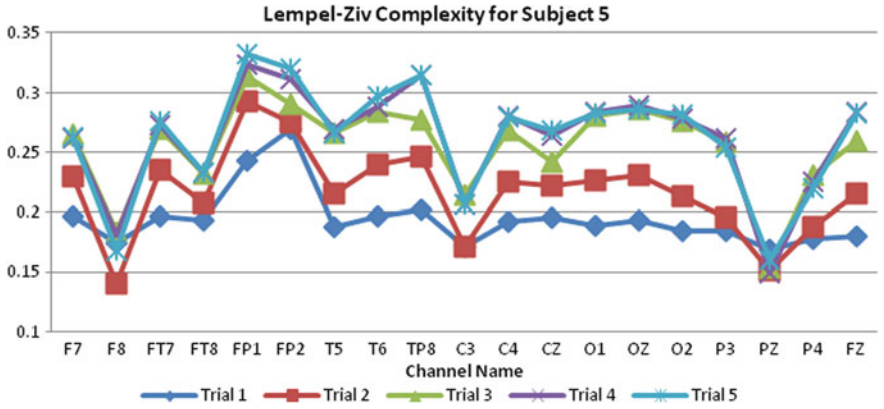


Fig. 9 Lempel-Ziv complexity distribution for all 19 EEG channels and all five trials of subject 5

4 Discussion

4.1 LZC Distribution

It would be a logical assumption that the since brain activity would decrease and be less complex as the task gets mastered, the LZC values would be shown to decrease with increasing number of trials performed. However, overall LZC values are shown to increase with increasing number of trials (see Fig. 11). Although this seems counter-intuitive since it shows that the signals get increasingly more complex even though the subject is getting more masterful with the task, one suggested possible explanation for this phenomenon is that at high mental

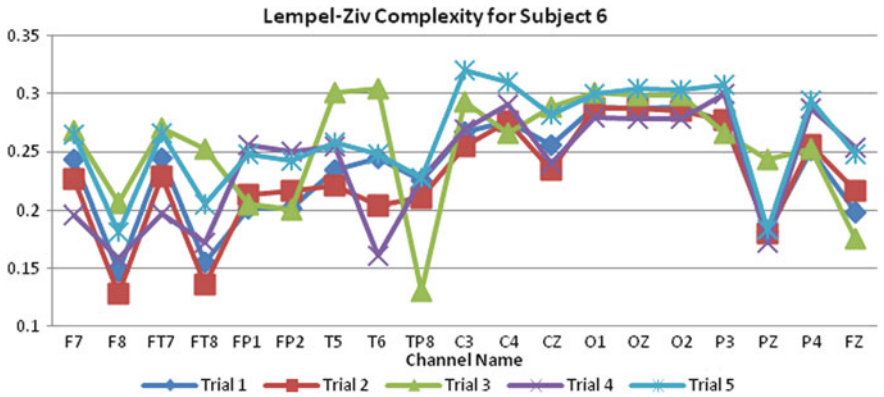


Fig. 10 Lempel-Ziv complexity distribution for all 19 EEG channels and all five trials of subject 6

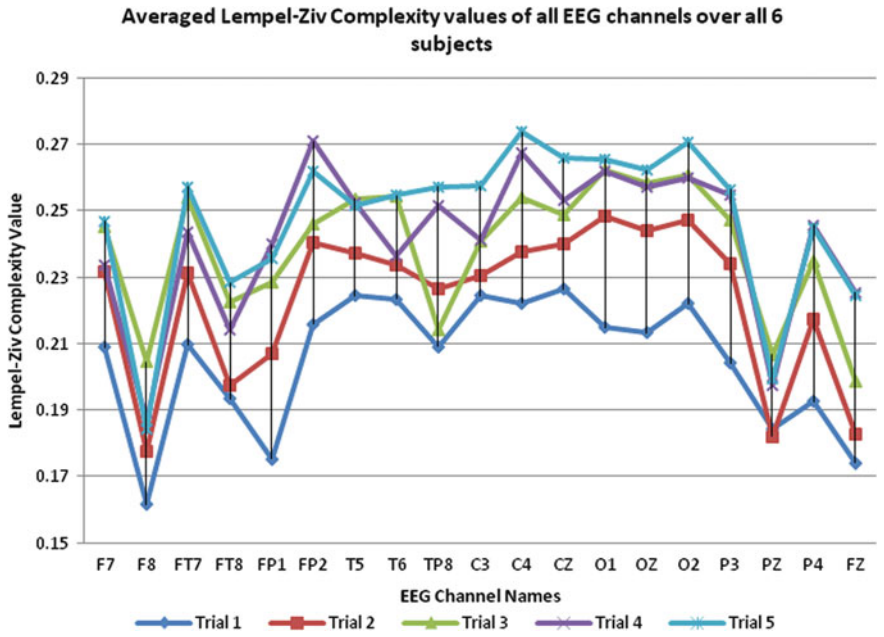


Fig. 11 Averaged Lempel-Ziv complexity distribution of all EEG channels over all 6 subjects

concentration, amplitude synchrony triggers the activation of an internal concentration state (Li et al. 2008; Hong et al. 2012). The internal concentration state refers to an isolation of the activity in the thalamencephalon cortex (Fernández et al. 1995). The result of that postulation is that at higher concentration with higher difficulty, the brain waves seemingly become more harmonic and synchronous, leading to lower LZC values. In our case, when the task gets mastered

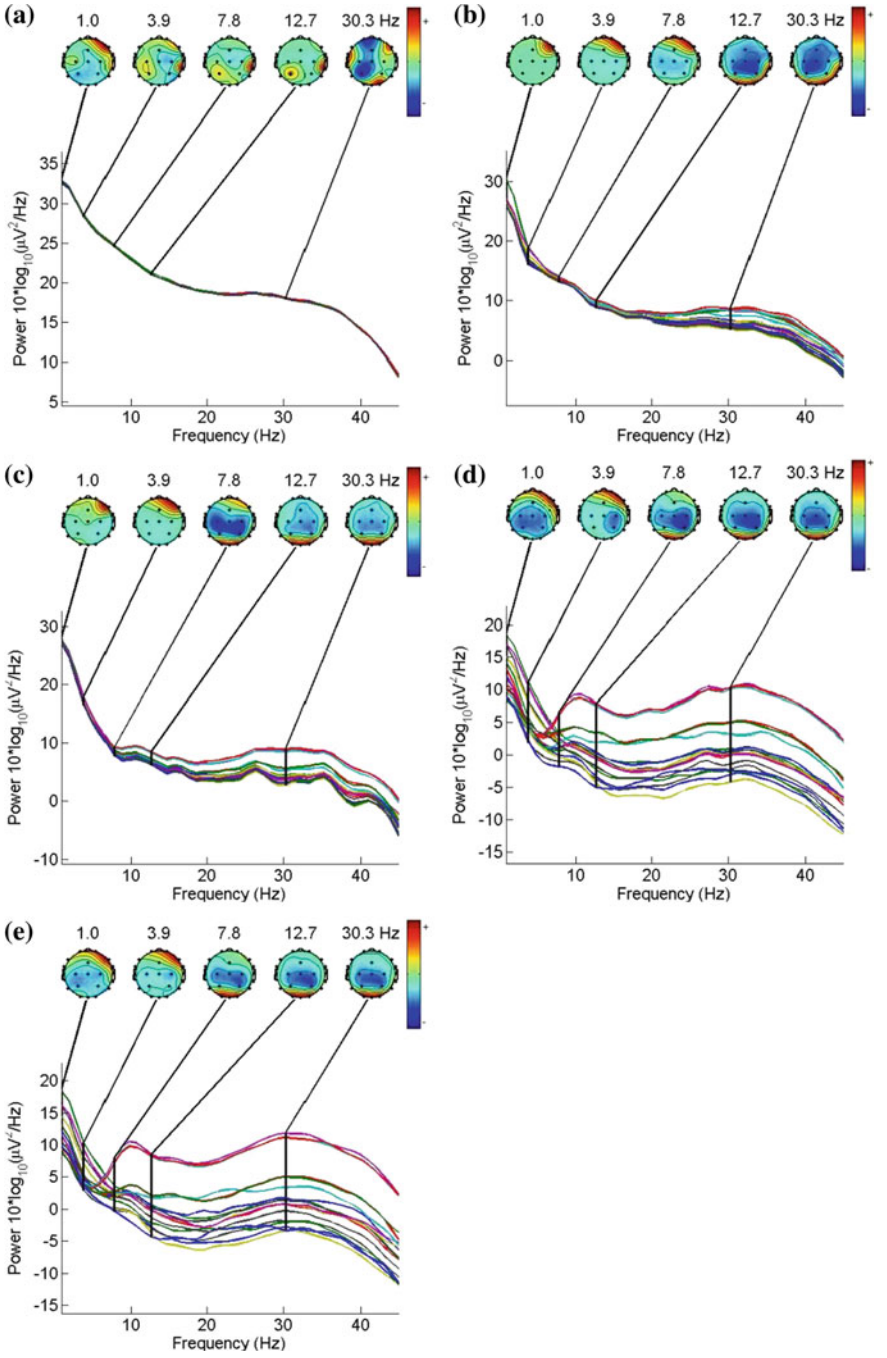


Fig. 12 **a–e** Channel spectra and topographic maps of subject 5 for all five trials. The figure parts are sequentially lettered from trial 1 to trial 5

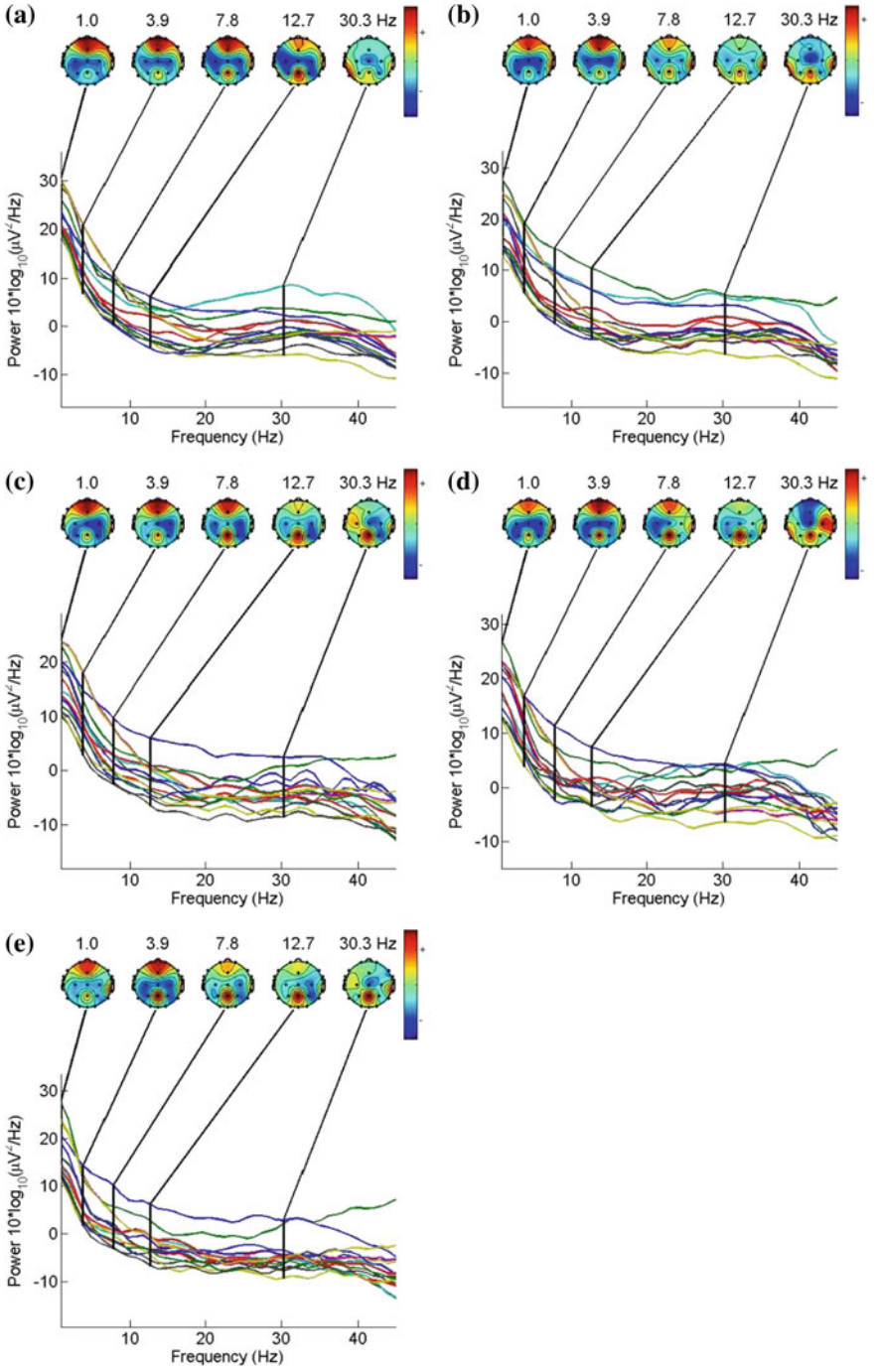


Fig. 13 **a–e** Channel spectra and topographic maps of subject 6 for all 5 trials. The figure parts are sequentially lettered from trial 1 to trial 5

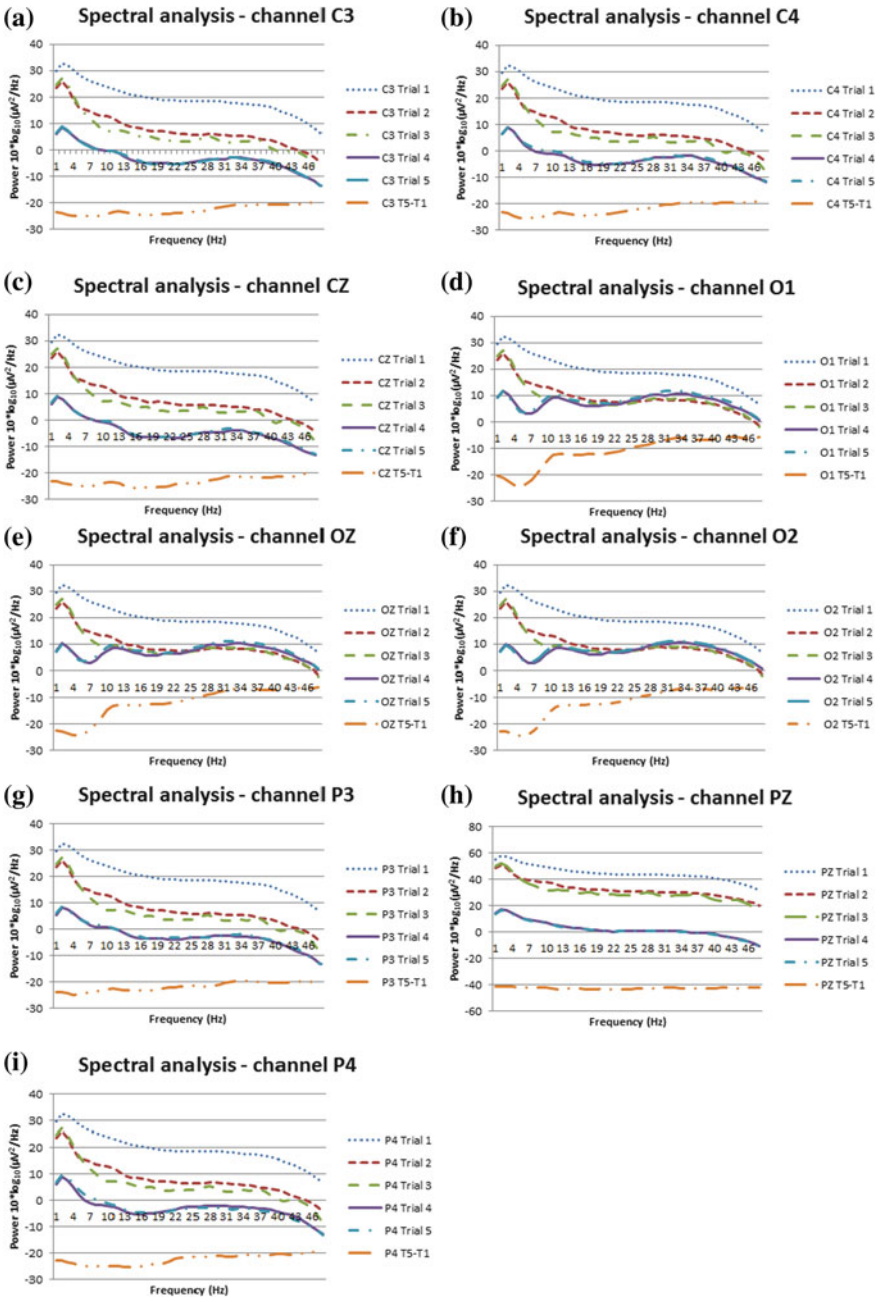


Fig. 14 **a–i** Individual frequency spectrums of channel C3, C4, CZ, O1, OZ, O2, P3, PZ and P4 for subject 5. The difference in power between trial 5 and trial 1 are plotted in orange for each of the channels

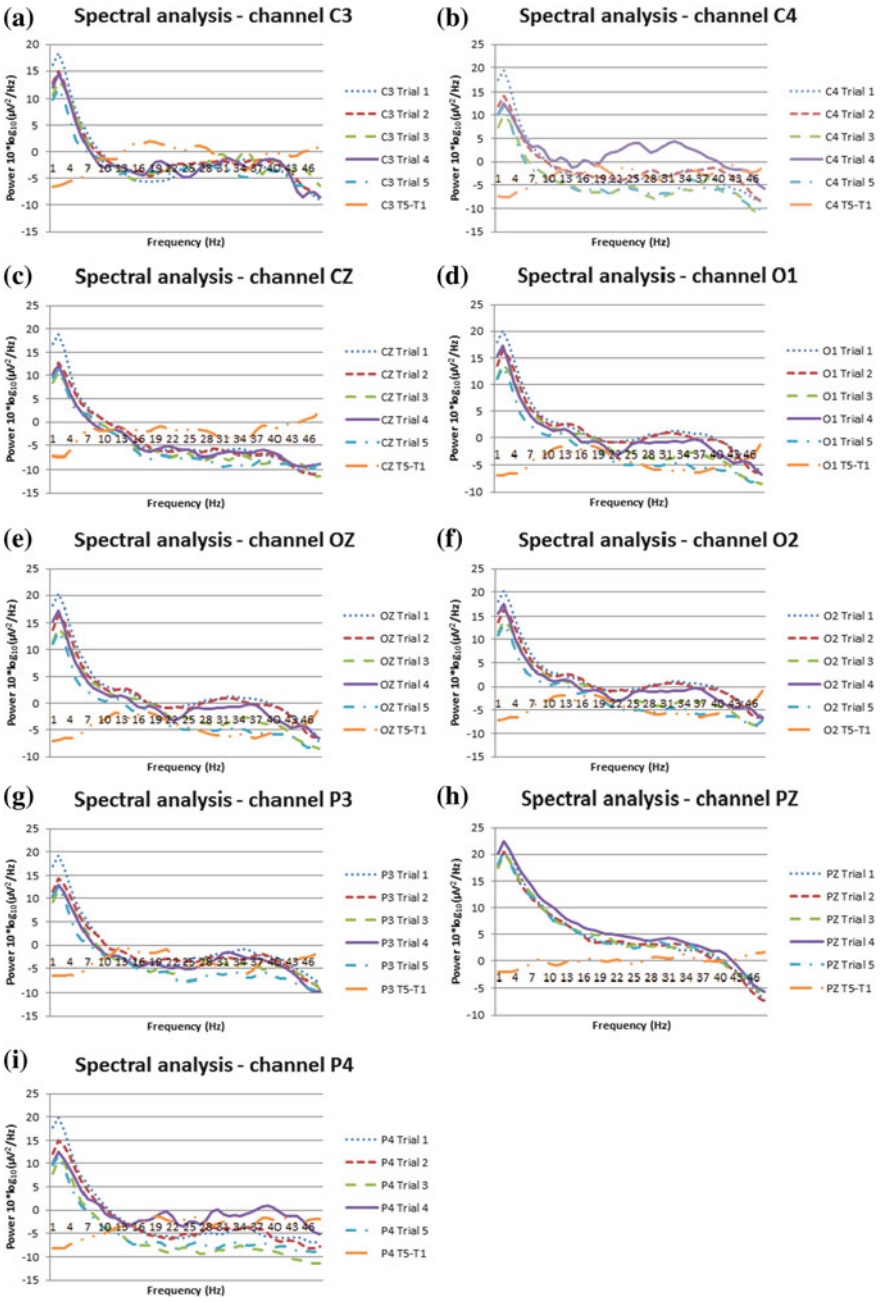


Fig. 15 **a–i** Individual frequency spectrums of channel C3, C4, CZ, O1, OZ, O2, P3, PZ and P4 for subject 6. The difference in power between trial 5 and trial 1 are plotted in orange for each of the channels

over increasing number of trials, the subjects need to concentrate less, thus the brain waves reduce to a more asynchronous and complex state.

By comparing the trends in LZC values of individuals (Figs. 5, 6, 7, 8, 9 and 10) over the five trials against the time taken for each trial, we can occasionally see that the increase in LZC values in some channels seem to mirror the improvement in timings. For example, when we look at subject 3, the time taken sharply improves significantly from trials 1 to 2 and similarly we see a large consistent rise in LZC values between trials 1 and 2. Similarly, for trial 2–5, the improvement in timings tapers off and a similar trend appears in the LZC values of trials 2–5. In addition, a similar corresponding trend is evident with subject 6 where subject 6 has relatively little improvement in timings and the LZC trend lines (see Fig. 10) are also shown to be clustered vertically. However, more studies need to be performed to determine the consistency of this trend as it might possibly be an artifact from the LZC computation due to the different lengths of the trials (Jing et al. 2006). Due to the rounding of data points, LZC has also been found to be more sensitive to high frequency components in the EEG signal (Ferenets et al. 2006).

Ideally, we should try to identify and isolate a consistently practicable trend in the LZC distributions that mirrors motor performance, meaning that the LZC trend lines of the 5 trials should not overlap each other and ideally be vertically spaced out to reflect the reduction in difficulty of consecutively performing the folding task over increasing number of trials. However, in most subjects, the LZC trend lines appear clustered together and intersecting at multiple channel locations. Visually, it seems hard to identify any useful LZC trend but when we average out the LZC values over all six subjects, as shown in Fig. 5.11, we can clearly see that channels C3, C4, CZ, O1, OZ, O2, P3, and P4 seem to satisfy our requirements for a useful LZC trend. This result mirrors that found by Hong et al. whereby their LZC distributions were significantly different at the occipital and parietal regions when comparing amongst different levels of mental workload using tasks of increasing levels of difficulty (Hong et al. 2012). In the following section, we shall compare these channels in the frequency spectrum to see if there is a similar trend present in these channels identified by their LZC distributions.

4.2 Spectral Analysis

Unlike the normalized LZC values, frequency spectrums can vary from person to person so we cannot directly average the values across the 6 subjects like in LZC analysis. By looking at the topological spectral analysis plots in Fig. 12a–e for subject 5, we can see that the delta (1–3 Hz) and theta (4–7 Hz) activity are frontally predominant and remain frontally predominant over the five trials. Similarly, the beta (14–25) and gamma (25–45 Hz) activity are predominantly occipital and persist over the five trials. However for the alpha band (8–13 Hz), it changes from being frontally dominant to occipital. Overall, we can see that the

power reduces across the whole spectrum in all the channels from trials 1–5, clearly indicating that brain activity and mental workload is reduced as the task is repeated.

By looking at the topological spectral analysis plots in Fig. 13a–e for subject 6, we can see similar topological trends as with subject 5 for the delta and theta range. However, the alpha range remains primarily frontal and parietal (corresponding to area around channel PZ). Similarly, the beta and gamma range are predominantly parietal instead of occipital as in subject 5. The overall activity level is also lower in subject 6 as compared to subject 5, especially in the first three trials. These two trends suggest that because subject 6 has better mastery of the task, he does not need to visually focus as much on the task. Hence, it is reflected in lower activity in the occipital region which processes visual information. Thus, the parietal region that is responsible for integration of visual and proprioceptive inputs (Medendorp et al. 2003) appears relatively more active hence stands out in contrast.

When we look at the individual channel frequency spectrums of subjects 5 and 6 (see Figs. 14a–i and 15a–i respectively), it is clear that with an increasing number of trials, the activity level of all the individual channels plotted drops. By looking at subject 5's channel plots (see Fig. 14a–i), we can see that the drop in activity in individual channels is more pronounced in the central channels (C3, C4 and CZ) and parietal channels (P3 and P4) as compared to the occipital channels (O1, O2 and OZ). This is also reflected in the overall channel spectrum trends as shown in Fig. 12a–e. In contrast, the overall activity levels of the individual channel frequency spectrum plots of subject 6 (see Fig. 15a–i) start out lower than that of subject 5 but do not drop as much as that in subject 5 from trials 1–5. This shows that subject 6's brain activity starts relatively low and remains low over the five trials. This trend is mirrored in the time taken for subject 6 to perform the five trials, suggesting that the task seemed naturally easier for subject 6 as compared to the other subject.

Research on the alpha band (8–13 Hz) (Klimesch et al. 2008) can be further divided into the slow alpha (8–10 Hz) (Klimesch 1999) and fast alpha (10–12 Hz) range (Klimesch et al. 1994, 1997a, b; Doppelmayr et al. 2002). Slow alpha while not topographically bound to any brain region, is shown to relate to general attention levels. Whereas theta band activity has been found to be closely related to short term memory load (Jensen and Tesche 2002; Raghavachari et al. 2001; Klimesch et al. 1999) and sustained attention (Ishihara and Yoshii 1972; Ishii et al. 1999; Kubota et al. 2001; Aftanas and Golocheikine 2001) In Fig. 14d–f we can see a sharp drop in the slow alpha and theta bands at the occipital area when we compare trial 1 with trial 5. However, the same trend is much more attenuated in subject 6 (see Fig. 15d–f). This result suggests that the slower subject 5 experienced higher levels of short term memory load and attentional demands during the early trial that decreases after more trials. This outcome complements the Fitts and Posner model for skill acquisition (Fitts and Posner 1967).

In the Fitts and Posner model, the first cognitive stage is associated with interpreting the nature of the task and trying out different strategies to perform the

task which result in highly variable performance and high cognitive load such as attention. The second stage is the associative stage where the person has settled on a strategy and the focus is on refinement of the strategy, biomechanical efficiency and task performance. The third stage is the autonomous stage where the person has achieved muscle memory and automation of the skill, enabling minimal cognitive load for the task, and allowing for the attention to be spent on other aspects such as higher order strategic planning. Therefore, classifying and monitoring the cognitive load of the trainee can be used as a method of determining the stage of skill learning where, according to the model, cognitive load and attention level decreases as the trainee advances in skill acquisition (Shumway-Cook and Woollacott 2007).

5 Conclusion

The objective of this study was to determine suitable means of gauging mental activity pertinent to the mastery of simple tasks. This paper presents our method of integrating EEG in serious games for mental patient's rehabilitation. We have shown in our results and discussion that LZC trends measured using EEG in the central, occipital, and parietal regions as well as spectral analysis of those channels, in particular within the theta and low alpha band, have potential to be used as markers for identifying the mental performance of a subject performing a simple training task. LZC and spectral analysis of EEG data is capable of revealing deeper mental processes that correlate with overall motor performance. These means of classifying cognitive activity enable alternative and more reliable ways of looking at a person's task mastery and hand eye coordination ability.

We have shown that using LZC and spectral analysis in combination can help to reduce the redundant processing of data in the EEG data that we recorded. In retrospect, by looking at the clustered plots of individual channels in Figs. 14 and 15, it is hard to visually identify the same trends found in LZC analysis. This shows that LZC and spectral analysis could be used simultaneously in order to get alternative perspectives on the same EEG data.

Using results from this study enables us to isolate the pertinent EEG markers needed to develop an effective solution through a Serious Game environment that uses both physical kinematics and mental states for the rehabilitation of mental patients. Combining human performance engineering with cognitive science in a computer gaming environment is our ongoing research for mental patient's rehabilitation.

References

Abásolo D, Hornero R, Gómez C, García M, López M (2006) Analysis of EEG background activity in Alzheimer's disease patients with Lempel–Ziv complexity and central tendency measure. Med Eng Phys 28(4):315–322. doi:10.1016/j.medengphy.2005.07.004

Aftanas LI, Golocheikine SA (2001) Human anterior and frontal midline theta and lower alpha reflect emotionally positive state and internalized attention: high-resolution EEG investigation of meditation. Neurosci Lett 310(1):57–60. doi:10.1016/s0304-3940(01)02094-8

Aftanas LI, Koshkarov VI, Pokrovskaja VL, Lotova NV, Mordvintsev YN (1996) Pre- and post-stimulus processes in affective task and event-related desynchronization (ERD): do they discriminate anxiety coping styles? Int J Psychophysiol 24(3):197–212. doi:10.1016/s0167-8760(96)00060-8

Chen X, Kwong S, Li M (2000) A compression algorithm for DNA sequences and its applications in genome comparison. Paper presented at the proceedings of the fourth annual international conference on computational molecular biology, Tokyo, Japan

Gusev VD, Nemytikova LA (1999) On the complexity measures of genetic sequences. Bioinformatics 15(12):994–999

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134(1):9–21. doi:10.1016/j.jneumeth.2003.10.009

Doppelmayr M, Klimesch W, Stadler W, Pöllhuber D, Heine C (2002) EEG alpha power and intelligence. Intell 30(3):289–302. doi:10.1016/s0160-2896(01)00101-5

Ferenets R, Tarmo L, Anier A, Jantti V, Melto S, Hovilehto S (2006) Comparison of entropy and complexity measures for the assessment of depth of sedation. IEEE Trans Biomed Eng 53(6):1067–1077. doi:10.1109/tbme.2006.873543

Fernández T, Harmony T, Rodríguez M, Bernal J, Silva J, Reyes A, Marosi E (1995) EEG activation patterns during the performance of tasks involving different components of mental calculation. Electroencephalogr Clin Neurophysiol 94(3):175–182. doi:10.1016/0013-4694(94)00262-j

Fitts PM, Posner MI (1967) Human Performance. Brooks Cole, Belmont

Friel PN (2007) EEG biofeedback in the treatment of attention deficit/hyperactivity disorder. Altern Med Rev 12(2):6

Hong J, Li X, Xu F, Jiang Y, Li X (2012) The mental workload judgement in visual cognition under multitask meter scheme. Int J Phys Sci 7(5):787–796. doi:10.5897/IJPS11.1372

Ishihara T, Yoshii N (1972) Multivariate analytic study of EEG and mental activity in Juvenile delinquents. Electroencephalogr Clin Neurophysiol 33(1):71–80. doi:10.1016/0013-4694(72)90026-0

Ishii R, Shinosaki K, Ukai S, Inouye T, Ishihara T, Yoshimine T, Hirabuki N, Asada H, Kihara T, Robinson SE, Takeda M (1999) Medial prefrontal cortex generates frontal midline theta rhythm. NeuroReport 10(4):675–679

Jensen O, Tesche CD (2002) Frontal theta activity in humans increases with memory load in a working memory task. Eur J Neurosci 15(8):1395–1399. doi:10.1046/j.1460-9568.2002.01975.x

Jing H, Jianbo G, Principe JC (2006) Analysis of biomedical signals by the Lempel-Ziv complexity: the effect of finite data size. IEEE Trans Biomed Eng 53(12):2606–2609

Klimesch W (1999) EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res Rev 29(2–3):169–195

Klimesch W, Doppelmayr M, Pachinger T, Ripper B (1997a) Brain oscillations and human memory: EEG correlates in the upper alpha and theta band. Neurosci Lett 238(1–2):9–12. doi:10.1016/s0304-3940(97)00771-4

Klimesch W, Doppelmayr M, Schimke H, Ripper B (1997b) Theta synchronization and alpha desynchronization in a memory task. Psychophysiology 34(2):169–176. doi:10.1111/j.1469-8986.1997.tb02128.x

Klimesch W, Doppelmayr M, Schwaiger J, Auinger P, Winkler T (1999) 'Paradoxical' alpha synchronization in a memory task. Cogn Brain Res 7(4):493–501

Klimesch W, Freunberger R, Sauseng P, Gruber W (2008) A short review of slow phase synchronization and memory: evidence for control processes in different memory systems? Brain Res 1235:31–44. doi:10.1016/j.brainres.2008.06.049

Klimesch W, Schimke H, Pfurtscheller G (1993) Alpha frequency, cognitive load and memory performance. Brain Topogr 5(3):241–251. doi:10.1007/bf01128991

Klimesch W, Schimke H, Schwaiger J (1994) Episodic and semantic memory: an analysis in the EEG theta and alpha band. Electroencephalogr Clin Neurophysiol 91(6):428–441. doi:10.1016/0013-4694(94)90164-3

Kubota Y, Sato W, Toichi M, Murai T, Okada T, Hayashi A, Sengoku A (2001) Frontal midline theta rhythm is correlated with cardiac autonomic activities during the performance of an attention demanding meditation procedure. Cogn Brain Res 11(2):281–287

Lempel A, Ziv J (1976) On the complexity of finite sequences. IEEE Trans Inf Theory 22(1):75–81. doi:10.1109/tit.1976.1055501

Li X, Hu B, Zhu T, Yan J, Zheng F (2009) Towards affective learning with an EEG feedback approach. Paper presented at the proceedings of the first ACM international workshop on multimedia technologies for distance learning, Beijing, China

Li Y, Tong S, Liu D, Gai Y, Wang X, Wang J, Qiu Y, Zhu Y Abnormal EEG complexity in patients with schizophrenia and depression Clinical Neurophysiol: Official J Int Fed Clin Neurophysiol 119(6):1232–1241

Medendorp WP, Goltz HC, Vilis T, Crawford JD (2003) Gaze-centered updating of visual space in human parietal cortex. J Neurosci 23(15):6209–6214

Monastra VJ, Lynn S, Linden M, Lubar JF, Gruzelier J, LaVaque TJ (2005) Electroencephalographic biofeedback in the treatment of attention-deficit/hyperactivity disorder. Appl Psychophysiol Biofeedback 30(2):95–114. doi:10.1007/s10484-005-4305-x

Nagarajan R (2002) Quantifying physiological data with Lempel-Ziv complexity-certain issues. IEEE Trans Biomed Eng 49(11):1371–1373. doi:10.1109/tbme.2002.804582

Raghavachari S, Kahana MJ, Rizzuto DS, Caplan JB, Kirschen MP, Bourgeois B, Madsen JR, Lisman JE (2001) Gating of human theta oscillations by a working memory task. J Neurosci 21(9):3175–3183

Shumway-Cook A, Woollacott MH (2007) Motor control : translating research into clinical practice, 3rd edn. Lippincott Williams & Wilkins, Philadelphia

Zhang L-Y, Zheng C-X (2009) Lempel-Ziv complexity changes and physiological mental fatigue level during different mental fatigue state with spontaneous EEG. Health 1(1):35

Zhang XS, Roy RJ, Jensen EW (2001) EEG complexity as a measure of depth of anesthesia for patients. IEEE Trans Biomed Eng 48(12):1424–1433. doi:10.1109/10.966601

Multi-link-ahead Conflicts Prediction in Dynamic Seaport Environments

Qing Li, Jasmine Siu Lee Lam and Henry Shing Leung Fan

Abstract Marine traffic conflict is an undesirable event of near misses between two moving vessels. Conflicts occur frequently in port waters and thus result in safety concerns as well as congestion and delays. A conflict between two vessels can be predicted via evaluating the relative positions of the vessels' domains. This chapter proposes an algorithm to predict likely conflicts multi-links-ahead before vessels actually encounter. A simulation model has been developed as a platform for implementation of conflict prediction in a dynamic traffic environment. An application of the model is demonstrated with the Port of Singapore. Simulation results show that an efficient and proper prediction would be two or three links ahead, and thus enables sufficient time for navigators to take evasive maneuvers.

Keywords Port traffic · Conflict · Conflict prediction · Vessel domain

1 Introduction

The Port of Singapore is one of the busiest transshipment ports in the world due to its geographical location, efficiency, and excellent connectivity. With substantial increases in marine traffic, the Port of Singapore is facing traffic congestion and potential risk of traffic incidents/accidents. Two major collision accidents were reported in the port waters of Singapore in 2009 and 2010, which caused severe damage to humans, assets, and the environment. In particular, the collision accident in May 2010 resulted in a serious crude oil slick near the east coast of Singapore.

Q. Li (✉) · J. S. L. Lam · H. S. L. Fan
School of Civil and Environmental Engineering, Nanyang Technological University, Nanyang, Singapore
e-mail: LiQing@ntu.edu.sg

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_5,

The causes of traffic incident/accident come down to a central issue: traffic conflict. Conflict refers to the situation of near misses between two moving vessels, which occurs frequently in seaports due to the special characteristics of port traffic as follows:

(1) Narrow fairways. Fairways are navigable waterways or channels which are open only to vessels with certain draught. Because of the limitations in geographical condition (e.g., width, depth, etc.), vessels cannot travel freely in fairways. Conflicts are prone to occur in a narrow fairway, where evasive maneuvers are limited due to insufficient space.
(2) High traffic density. Compared to the open sea, available space within a seaport is limited, but a larger number of vessels move in the traffic network. Port waters often have higher traffic density, especially during the peak period. This poses great potential risk of vessel conflicts.
(3) Complex traffic regulations. Port authorities establish a series of complex regulations for controlling and managing traffic. For example, according to geographical conditions, fairways are specified as one-way lane or two-way lane; vessels are assigned different priorities in operations either to give way or stand-by. Complex regulations need to be taken into consideration for a vessel to take corrective maneuvers in order to avoid conflict.

Compared with collisions, conflicts do not involve physical contact but relate to the situation of near misses. However, a conflict can also be considered the same as a collision to some extent. The risks resulting from collisions or conflicts only differ in their degree of severity in regards to navigational safety. Conflicts are general incidents, while collisions are dangerous accidents. Collisions present a kind of extreme cases in traffic conflicts (Debnath and Chin 2010; Weng et al. 2012). When a conflict cannot be properly resolved, it would lead to a collision accident which could cause a loss of life and property, and may even threaten the ocean environment.

Besides safety concerns, the most common result of a conflict is time delay which results from evasive maneuvers of vessels to avoid a collision with targets. As mentioned, the sea space of a busy seaport is finely meshed and intensively used due to increased marine traffic. Within a heavily loaded traffic network, even a small interaction may have a large impact on the entire network. Frequent delays in vessel operations would increase vessel-waiting time and the length of waiting queue, slow down the speed of vessel traffic in the network, and may finally result in traffic congestion.

We can see that a conflict is an undesirable event between vessels related to safety concerns as well as congestion and delay which affects the efficiency of port operations. Vessel conflict is a critical issue in marine traffic safety, and of great practical significance in traffic congestion management. For vessel encounters in the sea, taking evasive turns and/or speed adjustment is the most direct way to avoid a conflict. However, the effectiveness of evasive maneuvers depends on whether the risk of a possible conflict could be predicted accurately and timely. To enable effective conflict resolution, we should be able to predict potential conflicts and take corrective measures in advance.

Little research is done in the literature on conflict determination. Thus, we would review relevant studies in collision determination as an alternative to conflict. Two popular criteria are used in past studies for determining a collision risk: the closest point of approach (CPA) and ship domain.

The CPA criterion is applied with two parameters: distance of closest point of approach (D_{CPA}) and time of closest point of approach (T_{CPA}). The two CPA parameters indicate the collision risk between two vessels. The smaller values the higher risk of collisions. The CPA parameters are usually applied in a collision avoidance system to guide the vessel for proper anticollision maneuvers. The speed and/or course maneuver can be calculated according to the minimal D_{CPA} and T_{CPA} (Lenart 1999, 2000).

In restricted waters, such as narrow fairways, the CPA criterion is not applicable. Instead, ship domain has been proposed as a more comprehensive and accurate criterion. Ship domain can be explained as "a water area around a vessel which is needed to ensure the safety of navigation and to avoid collision" (Zhao et al. 1993). The first ship domain model for a narrow channel was proposed by Fujii and Tanaka (1971) based on the field observations. Later, Goodwin (1975) developed a domain model for open sea.

Ship domains proposed by various studies differ from one to another (Davis et al. 1980; Coldwell 1983; Zhu et al. 2001; Pietrzykowski 2008). Typically, the shape and size of a vessel domain depend on a number of factors (vessel's speed and length, sea area, traffic density etc.). In a port traffic system, vessels traveling along fairways are required to keep various safety clearances in accordance with the port's regulation. The domain of a vessel can thereby be referred to as the clearance area. We have implemented a simulation system to predict conflicts using the criterion of vessel domain. Before two vessels actually encounter, if the relative movement of one vessel's domain interferes with another vessel's domain, a potential conflict is predicted.

Previous study provides an algorithm for conflict detection through estimation of relative position between vessel domains (Li and Fan 2012). With this algorithm, we can detect a conflict likely occurs one-link-ahead current vessel position. As will be pointed out in the Sect. 2 the previous method assumes that the vessel will make a sharp turn at each node. But actually, a vessel will make a smooth turning at each node. Moreover, detecting only one-link-ahead is insufficient, particularly if the link is short. Therefore, this chapter proposes a new algorithm for extension of one-link-ahead prediction into multi-link-ahead prediction. With multi-link-ahead prediction, a possible conflict can be predicted more links ahead. The required number of links ahead is designed as a parameter in the simulation system. Such prediction ensures that the navigator has sufficient time to take actions before the predicted conflict occurs.

Simulation is an approach to model a real-life system on a computer so as to study how the system works. Simulation has good efficiency in integrating complex systems, such as the port traffic system concerned in this research; and good performance in computer animation, e.g., to mimic dynamic vessel movements and complex traffic scenarios. In addition, simulation is a useful adjunct or an

effective alternative to mathematical methods. A review of past studies revealed that traditional mathematical methods are quite complicated when used to estimate conflict risk (Zhu 2003). It is more feasible to develop a simulation model where data required can be substituted by parameters thus be simplified as basic input. This research proposes to develop a simulation system, called "Marine Traffic Conflict Simulation System", through which the function of conflict prediction will be implemented.

2 Simulation System Overview

2.1 Representation of a Seaport Traffic System

A seaport traffic system is a network of nodes and links. Within the network, each link indicates a fairway section, and a node can be

(1) a berthing/anchorage area,
(2) a boarding point for port pilots,
(3) an intersection area of fairways, or
(4) a separation point dividing a fairway into two sections due to differences in widths and/or traffic regulations.

Figure 1 shows an example of the Port of Singapore we use in the simulation model. The circular dots with different gray values represent different types of nodes: green dots refer to boarding points; blue dots refer to anchorage areas; red dots refer to berthing areas; and black dots refer to separation points and intersections of fairways. A rectangle connecting two nodes indicates a link. The width of a rectangle indicates the width of the link. Vessels are specified to travel along the link. Each vessel is visualized as a rectangle with a red arrow indicating the traveling direction.

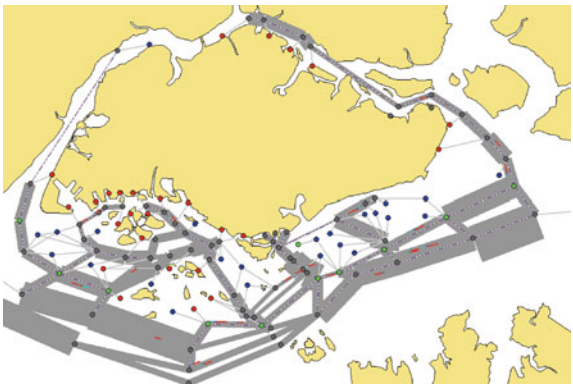


Fig. 1 A seaport traffic system for the port of Singapore

2.2 Notations for a Vessel and its Domain

A vessel is denoted as $V\left(\mathrm{O}, d, \Phi, \Psi, \bar{\Phi}, \bar{\Psi}^1, \bar{\Psi}^2\right)$. A vessel is simplified as a rectangle V centering at O (x, y) with dimensions Φ (width), Ψ (length), and d (traveling direction). The clearance area of a vessel is defined as a zone to keep enough distance to avoid conflicts with other vessels. The clearance area varies according to a vessel's outline, dimension, sailing speed, technical parameters and fairway characteristics. In our simulation system, the vessel's clearance area is a rectangle R. The lateral clearance is $\bar{\Phi}$. The longitudinal clearance is given by $\bar{\Psi}^1$ in the direction of the bow and $\bar{\Psi}^2$ in the direction of the stern. These parameters $\left(\bar{\Phi}, \bar{\Psi}^1, \bar{\Psi}^2\right)$ are set up as input data. Figure 2 shows a vessel with its domain.

2.3 Vessel Path Calculation

The simulation system requires the vessel path to be assigned so as to control vessels in which their positions can be tracked. This path refers to vessel trajectories based on basic maneuvers (except of particular actions, e.g., evasive turn for conflict avoidance). One path is a combination of trajectories in straight links and trajectories passing nodes.

It is supposed that a vessel keeps a straight line course in a link. Its trajectory is along the center line of traffic lane. As shown in Fig. 3a, red lines indicate vessel trajectories in a two-way lane link.

A vessel needs to make a turn to cross through an intersection or a junction of links. The vessel should keep a continuous and smooth moving during its crossing process. Constant radius turn technique is commonly used in marine navigation and piloting, for it enables a steady turn with less drift angle and less speed loss (Aarsaether and Moan 2007).

We proposed a method to determine vessel crossing trajectory based on constant radius turn maneuver. As is shown in Fig. 3b, a vessel is traveling along a link c_1 and moving toward to a link c_2. Regardless of the reflection time of rudder, vessel trajectory is a circle arc whose radius is determined by parameters of bend

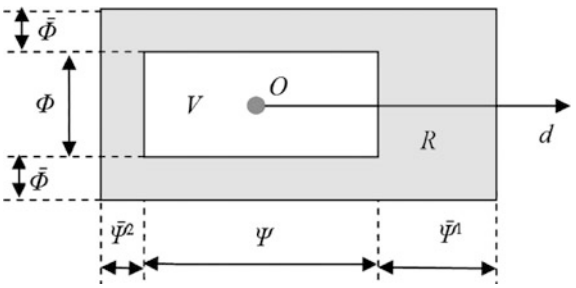


Fig. 2 A vessel and its domain

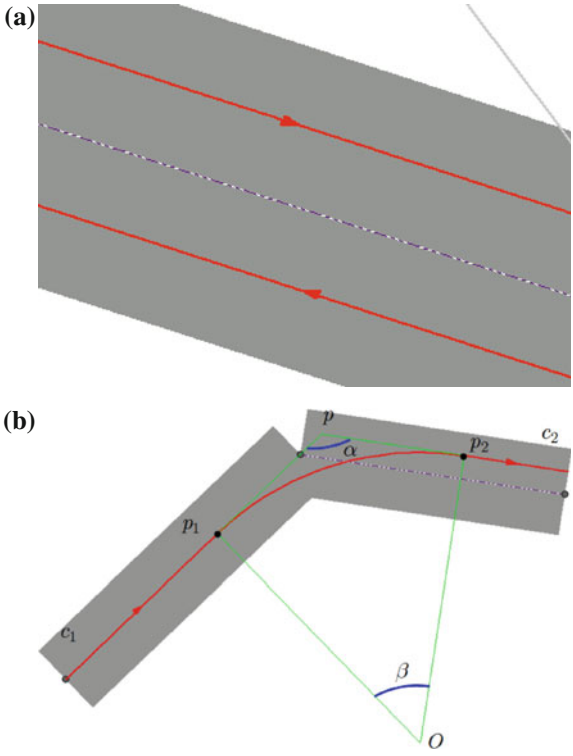


Fig. 3 Vessel trajectories **a** vessel trajectories in a link and **b** vessel trajectories when passing a cross

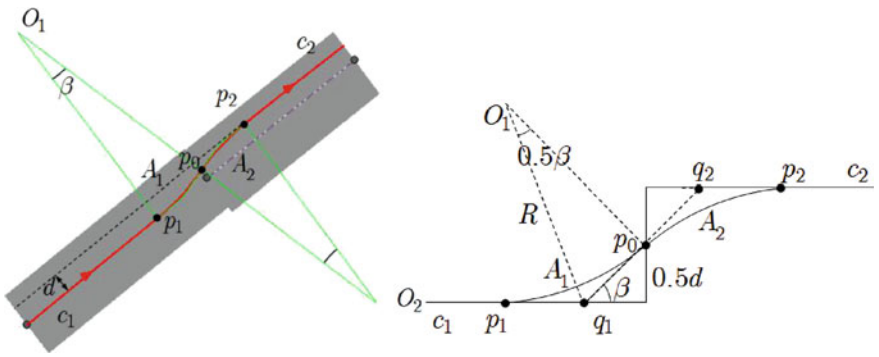
speed and rate of turn (ROT). The arc is tangent to both c_1 and c_2. The required circle arc is determined as long as the turning points p_1 and p_2 are obtained.

Suppose that the angle between the two links is α (degree). Bend speed and ROT are given as vessel parameters. We have the following equations,

$$\begin{aligned} \beta &= 180 - \alpha, \\ t &= \beta / r, \\ L &= 60vt, \\ R &= |op_1| = |op_2| = 360L/2\pi\beta, \\ D &= |p_1 p| = |p_2 p| = R.\tan(\beta/2). \end{aligned} \tag{1}$$

where

β: angle of the vessel needs to turn (degree),
t: time for the vessel makes the crossing,
L: length of the required arc,
v: bend speed,
R: radius of the required arc,
o: the center of the required arc,
p: intersection point of c_1 and c_2,



(a) **(b)**

Fig. 4 Vessel crossing trajectory in two parallel links (**a**) A vessel will cross through links c1 and c2, and (**b**) its trajectory is an S-shaped curve

p_1: turning point where the vessel starts turning,
p_2: turning point where the vessel ends turning, and
D: distance between p_1 to p (or p to p_2).

Sometimes a vessel needs to turn twice in order to cross through two parallel links. A similar method is used for determining vessel crossing trajectory. As shown in Fig. 4a, a vessel will cross through two parallel links c_1 and c_2. The vessel will make two turns during its crossing. Its trajectory is an S-shaped curve which is composed by two circle arcs, denoted as A_1 and A_2. It is supposed that p_1, p_2, and p_0 are turning points, which means that the vessel will make the first turn at p_1 to p_0, and then will make the second turn at p_0 to p_2. The turning points satisfy the following conditions:

(1) A_1 is tangent to c_1 at point p_1,
(2) A_2 is tangent to c_2 at point p_2, and
(3) A_1 is tangent to A_2 at point p_0.

To calculate the turning points, we assume that,

(1) The distance between c_1 and c_2 is d.
(2) The crossing trajectory is an S-shaped curve whose center is p_0 (Fig. 4 b). The angle of the vessel makes for each turn is β. Radius of each turn is R, which can be calculated with Eq. (1).
(3) q_1, q_2 are two points on c_1, c_2, which satisfy that line $\overline{q_1 p_0 q_2}$ is tangent to both A_1 and A_2 (Fig. 4 b).

Then we have the following equation,

$$|p_0 q_1| = d/(2\tan\beta) = R \cdot \tan(\beta/2) \tag{2}$$

Thus,

$$\beta = 2 \arcsin \sqrt{d/(4R)} \tag{3}$$

Since $L = |p_0q_1|$, we have the following equation,

$$L = |p_1q_1| = |p_0q_1| = |p_0q_2| = |p_2q_2| = R \cdot \tan(\beta/2) \tag{4}$$

The turning points of p_1, $p_{0,}$ and p_2 can be calculated with Eqs. (3, 4). The vessel's crossing trajectory is hereby determined.

In most cases, c_1 and c_2 are not so far apart that a vessel would turn for a small angle at each turn, i.e., β is less than 90°. It means

$$\beta = \arcsin \sqrt{d/(4R)} < \pi/4 \Rightarrow \sqrt{d/(4R)} < \sqrt{2/2} \Rightarrow d < 2R.$$

When c_1 and c_2 are so far apart, i.e., $d \geq 2R$, a vessel cannot cross through the two links even it turns for 90°. In this case, the vessel has to keep a straight path between the two turns. An example is shown in Fig. 5. The vessel's crossing trajectory is consisted of three parts: two circle arcs (i.e. A_1 and A_2) and a straight line ($\overline{r_1r_2}$). We have the following equations,

$$\begin{aligned} &\beta = \pi/4, \\ &|p_1o_1| = |o_1r_1| = |r_2o_2| = |o_2p_2| = R. \end{aligned} \tag{5}$$

where

p_1, r_1, $p_{2,}$ and r_2: turning points

O_1, O_2: the centers of A_1 and arc A_2.

With Eq. (5), we can calculate p_1, r_1, $p_{2,}$ and r_2. The vessel's crossing trajectory is thus determined (Fig. 5).

The path for a vessel consists of line segments and circular arcs, which are connected end-to-end. In this chapter, for the purpose of conflict prediction, each

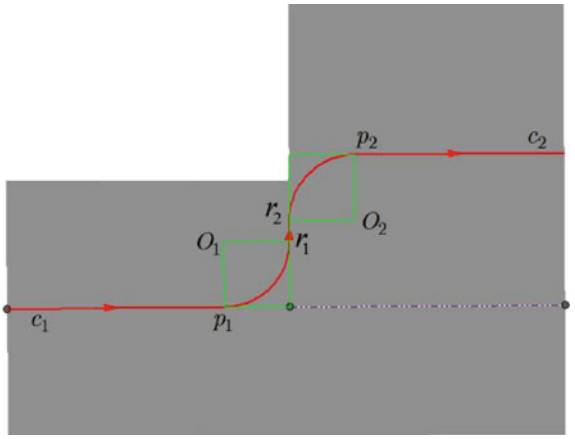


Fig. 5 Vessel crossing trajectory in a case of $D > 2R$

arc is approximated using a polygon by sampling points on the arc every 30^{o}. As a result, the path for a vessel on each link is a polygon. Each edge on the polygon is a section on the link, named sub-link. This is different from the previous work (Li and Fan 2012), where the path on each link is a single line segment. Consequence, the one-link-ahead prediction algorithm cannot work anymore.

2.4 System Design

This study aims to develop a simulation system, called "Marine Traffic Conflict Simulation System", which can run on advanced microcomputers or graphic workstations. Algorithm for conflict prediction outlined in Sect. 3 will be implemented through the simulation system. Design of the simulation system shall take the following considerations into account:

(1) It should be applicable in a variety of water areas;
(2) Complicated conflict scenarios (e.g., multivessel conflict) can be investigated;
(3) The process of simulation is displayed with dynamic graphics;
(4) A user interface which allows people to interact with simulation;
(5) Real-time data transmission and communication;
(6) Good compatibility and expansibility with other platforms;

Based on the above considerations, two operation modes may be provided in this simulation system: offline simulation and online simulation. Offline simulation is used to test and debug algorithms, as well as for planning and analysis; while online simulation is a platform used for real-time decision-making in real traffic situations. Offline simulation can be developed by individuals under laboratory environment. This makes implementation relatively easy. Compared to offline simulation, the implementation of online simulation depends on external cooperation for requirements of more hardware and real-time data.

3 Conflict Prediction

With given information on vessel characteristics, paths, and schedules, vessel movements within the traffic network can be displayed in a simulation system. Given a pair of vessels moving within the traffic network, we need to predict whether a potential conflict will occur between them at certain links ahead current vessel positions. The number of links ahead can be specified in the simulation system.

The prediction is performed at current time and estimate the conflict possibility until the meeting time of the two vessels. The time period for conflict prediction is divided into several time intervals. Conflict prediction is to evaluate the relative

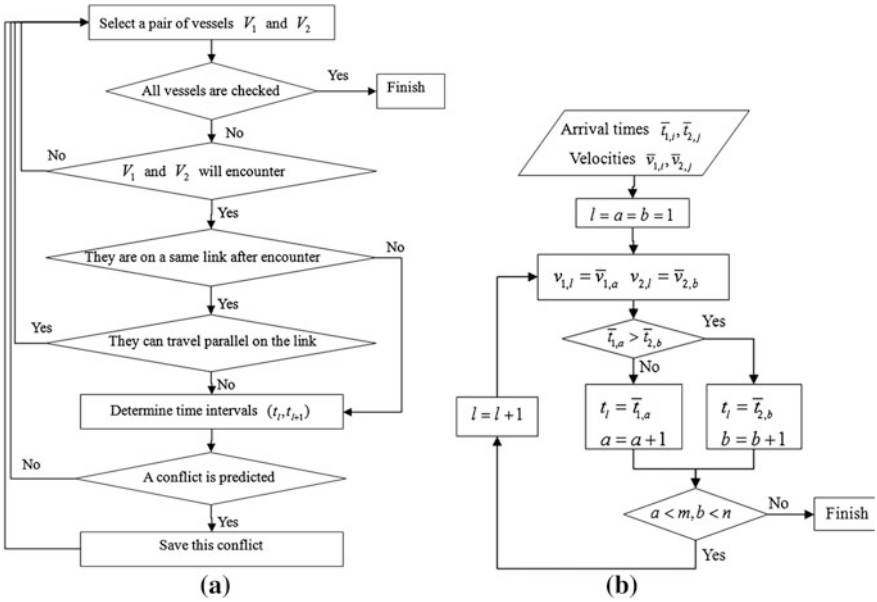


Fig. 6 Two flowcharts for conflict prediction **a** predict a potential conflict for any pair of vessels, and **b** divide the given time period for conflict prediction into small time intervals

movement of the domain of one vessel with respect to the domain of another vessel during each time interval. Figure 6 is the flowcharts for conflict prediction.

As shown in Fig. 7, there are three vessels V_1, V_2, and V_3 whose paths are represented by arrow lines. Potential conflicts may occur in each pair. Paths of V_1 and V_2 intersect at point A which is located near a node. A conflict is likely to occur when the two vessels cross through the node area. Likewise, the paths of V_2 and V_3 intersect at a point B in a link. In this case, the two vessels travel along a same link after point B, and may conflict in the link. Our method used to predict a conflict at a node or in a link is basically same. The only difference is that link width should be taken into account to predict a conflict in a link. If the width of a link is sufficient such that two vessels can travel in parallel, a conflict will not occur between them.

An example is used to describe the algorithm design for multi-link-ahead conflict prediction. In Fig. 8, suppose that conflict prediction is required to execute for n links (equivalently m sub-links) from current vessel positions onward. The first step of conflict prediction is to estimate whether a pair of vessels will encounter on those m sub-links. The intersection of two vessels' paths, e.g., point A and B in Fig. 7, are defined as the meeting point. According to the flowchart in Fig. 6a, if the meeting point exists between two vessels, the main steps for predicting a potential conflict are as follows:

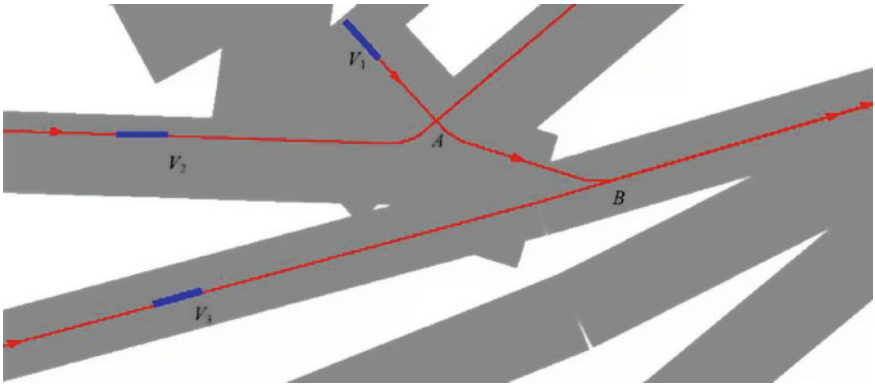


Fig. 7 Potential conflicts in traffic network

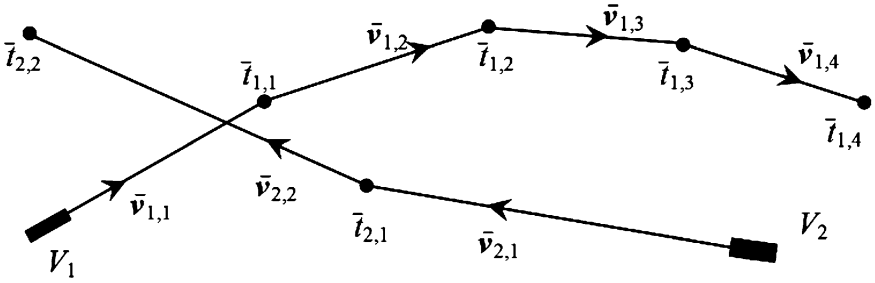


Fig. 8 Paths of two vessels

Step 1: obtain necessary vessel information, including arrival times at sub-links and speeds in each sub-link
Step 2: determine a set of time intervals such that in each interval, vessel speed keeps constant
Step 3: predict a conflict in each time interval.

Table 1 lists the necessary navigation information for V_1 and V_2 from their current positions until n more links forward. As shown in Table 1 and Fig. 8, the time when V_1 enters the ith sub-link is $\bar{t}_{1,i}$ and V_1 maintains a constant speed on the ith sub-link as $\bar{v}_{1,i}$. Similar navigation information can be calculated for V_2.

Table 1 Information of two vessels

Vessels	V_1	V_2
Number of sub-links ahead	m_1	m_2
Time to sub-links	$\bar{t}_{1,i}$	$\bar{t}_{2,j}$
Velocity on a sub-link	$\bar{v}_{1,i}$	$\bar{v}_{2,j}$

The information in Table 1 is collected in the time intervals $(0, \bar{t}_{1,m})$ and $(0, \bar{t}_{2,m})$ for V_1 and V_2, respectively. In Step 2, to predict conflicts during $\left(0, \min\{\bar{t}_{1,m}, \bar{t}_{2,m}\}\right)$, we divide the entire time period into a set of time intervals as (t_l, t_{l+1}), such that the speeds of V_1 and V_2 are constant as $v_{1,l}$ and $v_{2,l}$ in each interval, respectively. With the information in Table 1, the flowchart in Fig. 6b gives an approach to obtain all the time intervals. From the first time interval to the last time interval, Step 3 will check whether or not the two vessels conflict. Once a conflict is predicted in a certain time interval, the algorithm will save the conflict without checking the remaining time intervals.

In a time interval (t_l, t_{l+1}), we propose to predict the conflict using the relative movement of V_1 to V_2, which is the movement of the domain of V_1 with respect to the domain of V_2. Suppose

$w_1 = v_{1,l} - v_{2,l}$: the velocity of V_1 with respect to V_2,
$Q_{i,l} = \left(q^1_{i,l}, q^2_{i,l}, q^3_{i,l}, q^4_{i,l}\right)$: the domain of the vessel V_i at $t = t_l$,
$q^k_{i,l}$: the kth corner of the domain $Q_{i,l}$,
$p^{k+1}_{i,l}$ $p^k_{l,l}$: the kth edge of the domain $Q_{i,l}$.

As shown in Fig. 9, the relative movement of the corner $q^k{}_{1,l}$ to V_2 is a line segment $p^k{}_{1,l}q^k{}_{1,l}$ where

$$p^k_{1,l} = q^k_{1,l} + (t_{l+1} - t_1)w_l.$$

The relative movement of $q^k_i q^{k+1}_i$ to V_2 is a parallelogram $p^k{}_1 = q^k{}_{1,l}q^{k+1}{}_{1,l}p^{k+1}{}_{1,l}q^k{}_1$. If V_1 and V_2 conflict with each other, the movement of at least one edge of V_1 will intersect with the domain of V_2, i.e., $P^k_l \cdot Q_{2,1} \cdot \varnothing$. In summary, V_1 and V_2 will conflict in the time interval (t_l, t_{l+1}) if and only if the follow formula holds

$$\cup\left(P^k_l \cap Q_{2,1}\right) \neq \varnothing.$$

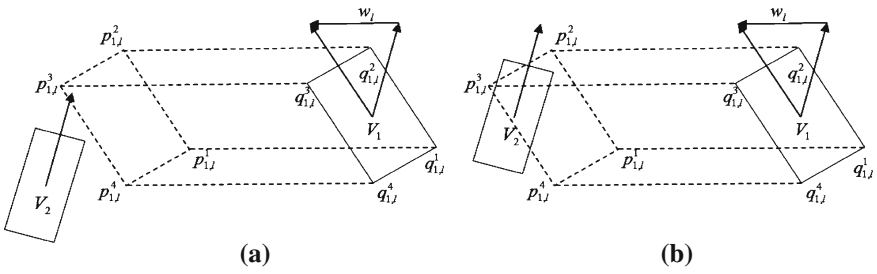


Fig. 9 Predicting the conflict in the time interval (t_l, t_{l+1}): **a** $P^k_l \cap Q_{2,l} = \varnothing$, V_1 and V_2 will not conflict with each other, and **b** $P^2_l \cap Q_{2,l} \neq \varnothing, P^3_l \cap Q_{2,l} \neq \varnothing$, V_1 and V_2 will conflict with each other

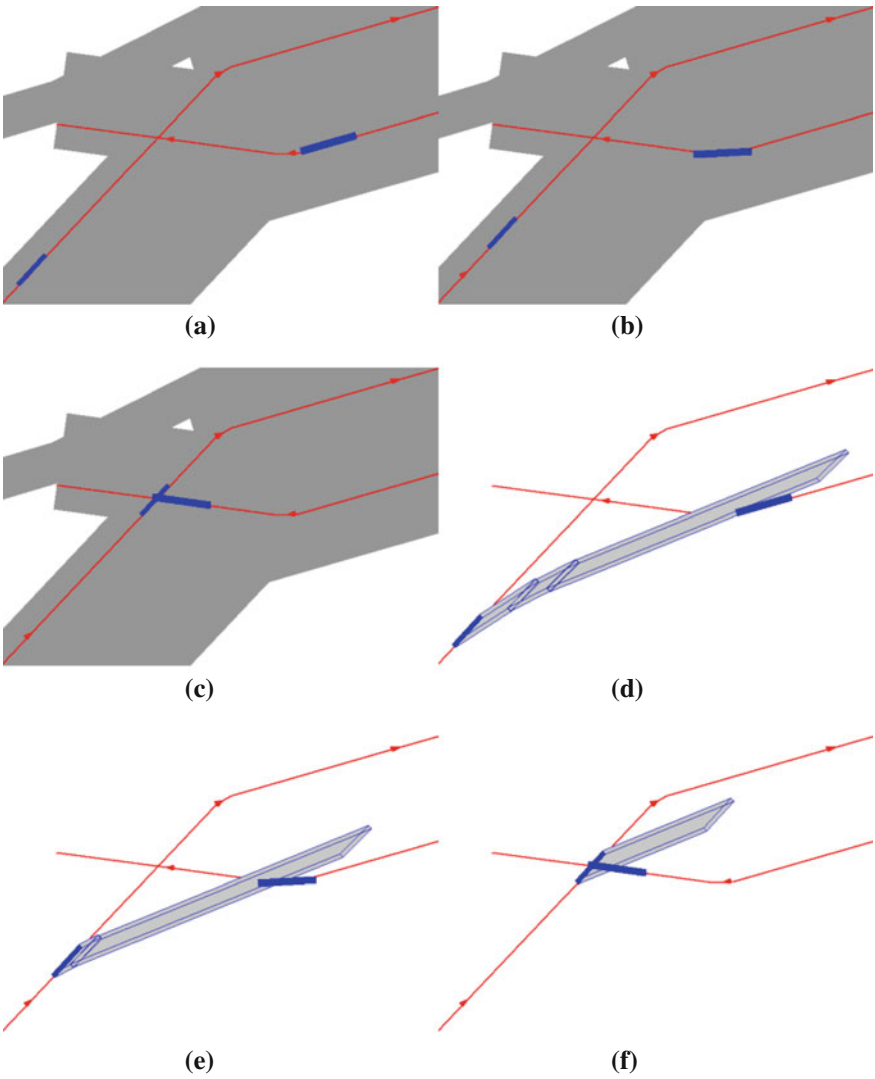


Fig. 10 Predicting the conflict at different locations **a–c** vessels at different locations, and **d–f** the relative movements at different locations

In this way, the conflict prediction is equivalent to check whether a parallelogram and a rectangle intersect or not. The example in Fig. 9a shows no conflict between V_1 and V_2 in the time interval (t_l, t_{l+1}) due to that $\cup\left(P_l^k \bigcap Q_{2,1}\right) = \varnothing$. However, from Fig. 9b, we have

$$P_l^1 \cap Q_{2,l} = \varnothing, P_l^2 \cap Q_{2,l} \cap \varnothing, P_l^3 \cap Q_{2,l} \cap \varnothing, P_l^4 \cap Q_{2,l} = \varnothing.$$

Therefore, there is a conflict between V_1 and V_2. In implementation, the algorithm starts from the first parallelogram $P_{1\ l}$ to the forth parallelogram $P4\ l$ to check the intersection $P_l^k \bigcap Q_{2,l}$. If one parallelogram intersects $Q_{2,l}$, a conflict is predicted and it is not needed to examine the remaining parallelograms.

4 Examples and Discussions

The algorithm of conflict prediction is implemented in our simulation system developed using Visual C++. Figure 10 gives a simple example for predicting a potential conflict between two vessels traveling toward to a node. Figure 10a–c shows changes of vessel movements from when they are far apart until they encounter. It is clear that a conflict occurs when they are crossing through a node (Fig. 10c). Correspondingly, the relative movements of vessel domains are shown in Fig. 10d–f. The relative movements are represented by the parallelograms enclosed by solid lines. In this example, two-link-ahead prediction is used, and time period for conflict prediction is divided into three intervals. The conflict is accurately predicted at the third time interval in Fig. 10d. Likewise, it is predicted at the second time interval in Fig. 10e.

An example of conflict prediction for multiple vessels is shown in Fig. 11, which contains 12 vessels. Four conflicts are predicted and listed in Table 2, which

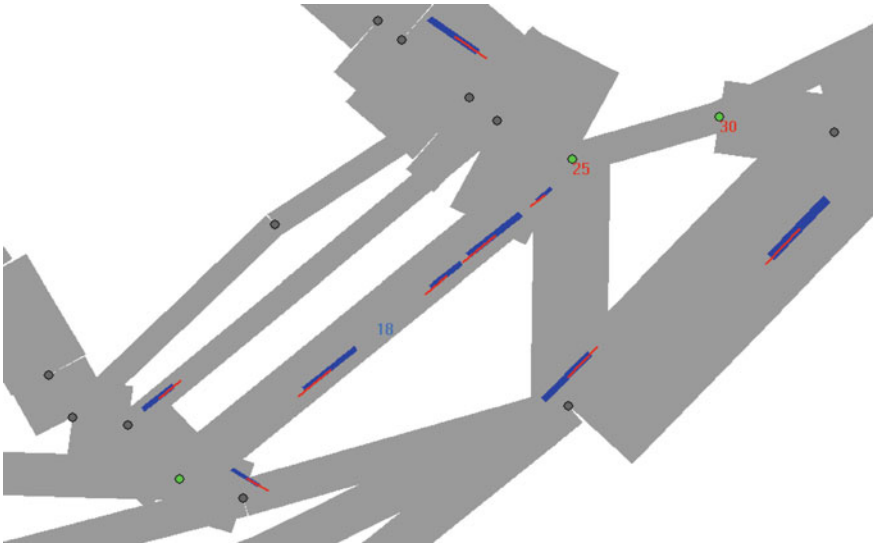


Fig. 11 An example of 12 vessels within the network (at current stage, there are only nine vessels; the number 18 indicates LINK18; the number 25 and 30 indicate NODE25 and NODE30)

Table 2 Statistics for conflict prediction

Vessels	Location	Conflict time[1]	Conflict prediction					
			$n = 2$		$n = 3$		$n = 4$	
			time[1]	second[2]	time[1]	second[2]	time[1]	second[2]
V_2 & V_{10}	LINK18	8:54:30	8:46:12	498	8:42:48	702	8:42:48	702
V_3 & V_{10}	LINK18	8:48:00	8:46:12	132	8:45:12	168	8:42:42	318
V_5 & V_8	NODE30	8:32:06	8:27:42	264	8:26:48	318	8:26:48	318
V_8 & V_{12}	NODE25	8:28:36	8:27:48	48	8:24:18	258	8:23:42	294

[1] The format for a time is hours : minutes : seconds

[2] the number of seconds that the conflict is predicted before the conflict time

also includes the result for n-links-ahead conflict prediction. With $n = 2$, in the worst case, the conflict prediction can predict the forth conflict 48 s in advance. It may not be enough for navigators to take safe actions to avoid the conflict. The problem can be solved by increasing the value for n. The conflict can be predicted 258 s in advance by increasing n to 3, and 294 s in advance by increasing n to 4.

5 Conclusions

A new conflict prediction algorithm has been proposed and implemented. The algorithm is designed to predict the potential conflict by checking the relative movement between two moving vessels. The algorithm simplifies the conflict prediction problem as an estimation of whether a parallelogram intersects with a rectangle conflicts in fairways, junctions and intersections would be predicted long time before the encounter of vessels. It enables that operators have enough time to take actions to avoid the conflict. Simulation results show that the algorithm is efficient. The logic of conflict prediction is applicable to other traffic systems by changing the input data. The simulation model is a generic model which can be adapted to other busy seaports that are faced with traffic congestion and delays. One future work is to improve the compatibility of the simulation system, so that it can be adopted in more complicated scenarios.

References

Aarsaether KG, Moan T (2007) Combined maneuvering analysis, AIS and full–mission simulation. Int J Mar Navig Saf Sea Transp 1(1):31–36

Coldwell TG (1983) Marine traffic behaviour in restricted waters. J Navig 36(3):430–444

Davis PV et al (1980) A computer simulation of marine traffic using domains and arenas. J Navig 33(2):215–222

Debnath AK, Chin HC (2010) Navigational traffic conflict technique: a proactive approach to quantitative measurement of collision risks in port waters. J Navig 63(1):137–152

Fujii Y, Tanaka K (1971) Traffic capacity. J Navig 24(4):543–552
Goodwin EM (1975) A statistical study of ship domains. J Navig 28(3):328–344
Lenart AS (1999) Manoeuvring to required approach parameters-CPA distance and time. Ann Navig 1:99–108
Lenart AS (2000) Manoeuvring to required approach parameters-distance and time abeam. Ann Navig 2:81–88
Li Q, Fan HSL (2012) A simulation model for detecting vessel conflicts within a seaport. Int J Mar Navig Saf Sea Transp 6(1):11–17
Pietrzykowski Z (2008) Ship's fuzzy domain-a criterion for navigational safety in narrow fairways. J Navig 61(3):499–514
Weng J, Meng Q, Qu X (2012) Vessel collision frequency estimation in the singapore strait. J Navig 65(2):207–221
Zhao J et al (1993) Comments of ship domains. J Navig 46(3):422–436
Zhu JQ (2003) Probabilistic conflict risk model for marine traffic within a sea port. Master thesis, Nanyang Technological University
Zhu XL et al (2001) Domain and its model based on neural networks. J Navig 54(1):97–103

Long Vehicle Turning

Yong Chen, Eng Sing Chua, Daniel Thalmann, Yiyu Cai, Yi Gong, Teng Sam Lim and Peng Wong

Abstract Safety and efficiency are two major issues when it comes to long vehicle driving and operating, particularly those with very large lateral and longitudinal sizes. Trajectory planning for long vehicle tuning always plays an important role to secure safety and efficient operation, especially when turning in narrow spaces limited by surrounding objects such as buildings. This chapter discusses a trajectory calculation method for long vehicle turning, based on a set of differential equations. The solution can be numerically obtained for any trajectory under different circumstances. We develop a generalized and systematic mathematical approach to determine the trajectories swept by each wheel and other related components of the vehicle. The envelope of the trajectories of the vehicle can then be derived according to geometric relationships and characteristics. Based on numerical analysis results, a 3D simulation is developed in this work for different types of long vehicles along with different given turning roads surrounded by buildings and other objects. This way we are able to do trajectory planning for long vehicle turning.

Keywords Vehicle turning · Simulation · Trajectory planning

Y. Chen · E. S. Chua · D. Thalmann · Y. Cai (✉)
Nanyang Technological University, Nanyang Avenue, Singapore
e-mail: myycai@ntu.edu.sg

Y. Gong · T. S. Lim · P. Wong
PEC Limited, Shipyard Road, Jurong, Singapore

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_6,

1 Introduction

1.1 Background

Safety and efficiency issues of long vehicle turning can be attributed to a number of factors, including vehicle shape, its control system, driving skills of the operators, roads, and nearby environment. Long vehicles are difficult to maneuver and sometimes are even dangerous especially in a complex environment due to the space limitation. For example, at intersections, roundabouts, or bus terminals, the accessibility and safety need to be analyzed and evaluated before maneuvering the long vehicles. It is not easy to have long vehicles changing lanes, reversing, or parking. To secure a safety driving, planning the collision-free turning for long vehicles under different conditions is necessary and important. The procedures of analysis and evaluation include:

(1) Compute the path swept by the long vehicles along the proposed trajectory;
(2) Determine whether there are intersections between the swept area and the surrounding obstacles;
(3) Modify the pre-planned path until no intersections are found in procedure (2).

To provide a mathematical description of the path swept by a moving vehicle, the maximal displacement of the rear end of the bus is analyzed for studying the driving hazard issues (Baylis 1973, Bender 1979). After that, the solution to the motion equation of midpoint of the rear axle is investigated, and then applied to different types of motions including straight line motion, circular motion, and others. Also, several applications such as turning around a corner, changing lanes, and traffic circles are discussed based on the calculated results (Freedman and Riemenschneider 1983). The limitation of these methods is that the implicit solution is obtained with the assumption that the motion of the front of the bus can be described as a solution of the second order linear differential equation. Numerical computing and analysis are introduced to simulate the maneuvering situations using several simplified models of truck braking and handling (Fancher and Balderas, 1987). The models involving vehicle dynamics concepts are used for preliminary analyses of off-tracking, straight line braking, steady turn, etc. In addition to normal vehicles, off-tracking of multiple unit vehicle combinations is discussed using a bicycle wheel model (Sayers 1986, Erkert et al. 1989). Mathematical methods are increasingly used for vehicle path planning and navigation. For purpose of autonomous path-tracking, the curvature that will drive the vehicle to the goal point is calculated geometrically with tracking algorithms such as pure pursuit and its variations (Coulter 1992, Hellström and Ringdahl 2005). Based on the model adapted from J. Baylis, the maximum rightward displacement of the vehicle depends on the relationship between the turning angle of the front wheel and the angle between the vehicle and the direction of the roadway (Bender 2000).

On the other hand, interactions between vehicles and roadway geometry are summarized and discussed from the perspectives of the comprehensive truck and roadway separately (Battelle Team 1995). Such studies focus more on the vehicle performance characteristics such as low-speed off-tracking, rather than investigating the whole scenario of the trajectories swept by the vehicles. To address the issue of the track of a bicycle back tire, the matrix Riccati equations are derived with the assumption that the tangent vector to the path of the rear tire always points to the contact point on the path of the front tire (Dunbar et al. 2001). To provide a close to reality simulation, the behavior of the vehicle turning is always analyzed in terms of kinematics and kinetics (Sweatman et al. 1991, Farmer 2008). However, it needs accurate and detailed kinematic parameters and related sensors are required. Also, some efforts have been made by applying intuitive techniques such as representing the transition curves with polynomials (Fioretti et al. 2008). It does not reflect the actual motion of the vehicle, although the computation is time efficient. Therefore, developing a generalized mathematical representation and solution will be necessary and effective, especially for those low-speed vehicles.

1.2 Objective

This chapter focuses on the trajectory calculation for long vehicles. To make trajectory planning, we propose a systematic and applicable approach for long vehicle turning. We attempt to develop a mathematical method for trajectory calculation without considering driving speed and angular turning velocity. The calculation results will be used for a 3D simulation of maneuvering long vehicles such as turning and parking in relation to the given surrounding buildings and other objects.

2 Mathematical Modeling

We start from a four-wheel long vehicle with an Ackermann steering geometry, which can be represented as a box (Fig. 1). Throughout the whole calculation and the simulation process, all of the wheels are assumed to have no slippage or skidding, and contact area between the tire and the ground is zero.

Based on the assumptions above and the fact that the rear axle is fixed on the vehicle body and each wheel rotating around the rear axle cannot change the direction on its own, it is almost as if there is an imaginary steered wheel located at point P (the midpoint of the front axle) with point Q (the midpoint of the rear axle) pointing to it. Thus the path of point Q and the line of centers PQ should be tangent to each other at point Q. φ denotes the angle between PQ and the vertical line while θ is the angle between PQ and front wheels (turning angle). With the

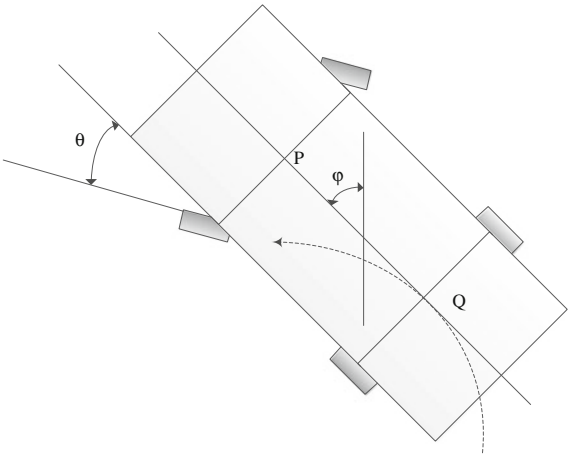


Fig. 1 Bottom view of a long vehicle with four wheels

assumption of no sideslip, there are three principles the steering process has to follow:

(1) The distance between midpoint of front axle P and midpoint of rear axle Q keeps constant and equal to the wheelbase;
(2) The direction of the velocity of midpoint of rear axle Q should point toward the midpoint of front axle P;
(3) The direction of the velocity of Q is always tangent to its trajectory.

2.1 Stable Circular Movement

Let us start with a simple question: how do we know the collision occurred between the vehicle and the walls (or other objects)? A straightforward solution is to find all the trajectories formed by the vehicle outline sweeping along the path; i.e., the trajectory profile or envelope of trajectories. If the envelopes of the trajectories and surrounding objects have no overlaps (intersections), the safety can be ensured for the vehicle to turn around the corner following the given path. Trajectories of front wheels are easy to get because these two wheels are totally controllable, consequently trajectories of these two points can be conveniently changed or modified with the aid of control system of the vehicle. The remaining problem is how to determine the trajectories of rear wheels with known parameters and conditions.

First, consider a simple case in which the vehicle is driven on a circular road, i.e., the front wheels are moving along a circular path. From common experience we know that the trajectories of rear wheels are also in the form of a circle.

To verify, a mathematical model (Prince and Dubois 2009) can be used for describing and solving this problem.

Vector-valued functions which depend on a single parameter t, representing time, can be used to determine the positions of P and Q, respectively:

$$\begin{aligned}\dot{\vec{Q}}(\mathrm{t}) &= \gamma(t)\left(\vec{P}(t) - \vec{Q}(t)\right)\\ \vec{B}(t) &= \vec{Q}(t) - \vec{P}(t)\end{aligned} \tag{1}$$

The scalar function $\gamma(t)$ represents the speed of point Q. Also, ψ is the angle between the horizontal line and the vector PQ, and it is equivalent to $\frac{\pi}{2} - \varphi$ in the case shown in Fig. 1. The Length of PQ is denoted as L while coordinates of point P are (P_1, P_2).

The general equation can be obtained:

$$\frac{d\psi}{d\mathrm{t}} = \frac{\sin(\psi)\dot{P}_1 - \cos(\psi)\dot{P}_2}{L} \tag{2}$$

For a circular path with radius R, we write it in parametric form and substitute it into Eq. (2), a differential equation with initial conditions can be obtained:

$$\begin{aligned}\vec{P}(t) &= R(\cos(\Omega(t)), \sin(\Omega(t)))\\ \frac{d\psi}{dt} &= -\dot{\Omega}\frac{R}{L}\cos(\psi - \Omega)\\ \psi(0) &= \psi_0,\ \Omega(0) = \Omega_0\end{aligned} \tag{3}$$

Actually, the equation above has a constant solution without applying the initial conditions, which shows that the trajectory of point Q is a circle of radius $\sqrt{R^2 - L^2}$ (if $R > L$) centered at the origin:

$$\cos(\psi - \Omega) = -\frac{L}{R} \tag{4}$$

2.2 General Form of the Trajectory

However, Eq. (4) indicates the final (stable) motion status of the vehicle, and it does not deliver any information about trajectory of Q at the very beginning when starting from the original stationary state to gradually merge onto the circular trajectory. We are interested in the locus curve of Q when the initial position of Q is not right on the inner circle. When steering around a corner, the movement of point P consists of several stages, as illustrated in Fig. 2, which shows the process of a long vehicle turning around a corner with 90°. R_P and R_Q represent the radii of the trajectories of P and Q, respectively, with the common center O. The initial position of vehicle body PQ is indicated in bold, while $P'Q'$ is its final position.

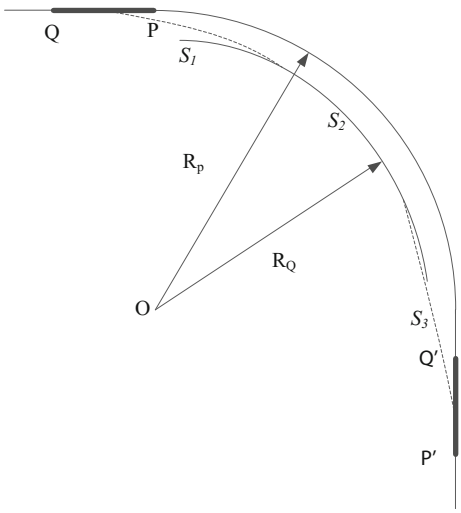


Fig. 2 Steering around a corner with 90°

(1) S_1: The dashed line S_1 denotes the trajectory of Q when P enters the circular path with radius R_P from the straight line motion. Although P can exactly follow the predefined trajectory, the path of Q is not a circle because Q cannot change its position from the straight line to the inner circle immediately.
(2) S_2: When Q enters its circular path, the vehicle body constrains Q along the inner circle with radius R_Q just by keeping P moving along the outer circle and this process has been illustrated in Eq. (4).
(3) S_3: The dashed line S_3 indicates the process when the vehicle begins to exit the circular arc. The trajectory of Q changes from a circular arc to a straight line gradually.

With S_2 easy to be calculated, S_1 and S_3 remain to be solved. A possible way to solve this problem is to obtain the trajectory of point Q depending on the velocity and acceleration of point P, by assuming that the motion of the vehicle can be represented as a solution of a second order linear differential equation, to obtain the trajectory of point Q depending on the velocity and accelerate of point P. However, it is not always applicable to set up such a differential equation sometimes. Here, following the three principles mentioned before, we attempt to determine the trajectory of midpoint of rear axle irrespective of kinetic and kinematic parameters. Concentrating on the components those play major roles in the motion while ignoring other irrelevant mechanisms, we can get a simplified model as shown in Fig. 3. Every point has its own behavior which should obey all the three principles. Point P can move freely in any direction while the locus curve of Q is determined by P: at any point on the curve, the tangent vector points to a corresponding point on the curve formed by P, and the distance between these two points is fixed.

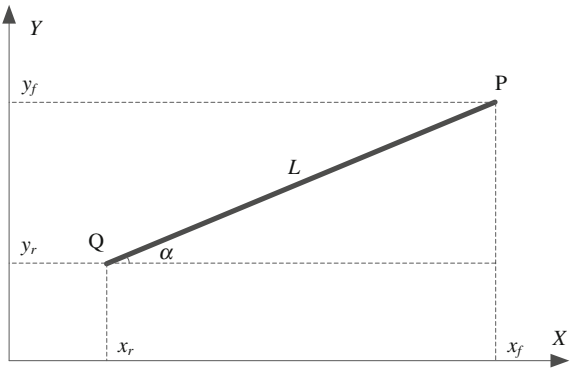


Fig. 3 Relationship between P and Q

If we express the trajectory coordinates of P and Q in parametric form, we can obtain the unit vector of velocity at point Q (serves as an intermediate variable):

$$\frac{\vec{v}}{\|\vec{v}\|} = (\cos\,\alpha,\,\sin\,\alpha) = \left(\frac{x_{\mathrm{r}}{}'}{\|\vec{v}\|}, \frac{y_{\mathrm{r}}{}'}{\|\vec{v}\|}\right) = \left(\frac{x_{\mathrm{f}} - x_{\mathrm{r}}}{L}, \frac{y_{\mathrm{f}} - y_{\mathrm{r}}}{L}\right) \tag{5}$$

According to the first principle, the length L keeps constant so that PQ can be considered as a rigid rod. The constraint imposed here is that the components of the velocities along the rod at these two points should be equal. Note that the resultant velocity at point Q is also along the rod. Hence,

$$\|\vec{v}\| = x_{\mathrm{f}}{}'\cos\,\alpha + y_{\mathrm{f}}{}'\sin\,\alpha \tag{6}$$

Replacing (6) in (5), we have:

$$\begin{aligned} x_{\mathrm{r}}{}' &= \frac{1}{L^2}\left[x_{\mathrm{f}}{}'(x_{\mathrm{f}} - x_{\mathrm{r}})^2 + y_{\mathrm{f}}{}'(x_{\mathrm{f}} - x_{\mathrm{r}})(y_{\mathrm{f}} - y_{\mathrm{r}})\right] \\ y_{\mathrm{r}}{}' &= \frac{1}{L^2}\left[x_{\mathrm{f}}{}'(x_{\mathrm{f}} - x_{\mathrm{r}})(y_{\mathrm{f}} - y_{\mathrm{r}}) + y_{\mathrm{f}}{}'(y_{\mathrm{f}} - y_{\mathrm{r}})^2\right] \end{aligned} \tag{7}$$

Because there is a possibility that the vehicle body is so long that the stage two does not exist at all for a 90° corner. In this case, stage three takes place after stage one or even during stage one. To make the process of stage one clear, we can investigate a complete circle. Then, the midpoint of the front axle is set to move along the circle with radius R, as shown in Fig. 4 below.

The parametric equation of P and initial conditions are as follows:

$$\begin{aligned} x_{\mathrm{f}}(t) &= R\,\sin(\omega t) \\ y_{\mathrm{f}}(t) &= R\,\cos(\omega t) - R \\ x_{\mathrm{r}}(0) &= -L,\; y_r(0) = 0 \end{aligned} \tag{8}$$

Because the equation set (7) is difficult to solve analytically, numerical method is an alternative way with the help of Matlab. Figure 5 illustrates the motion at three different times, and it shows how trajectory changes from linear motion to

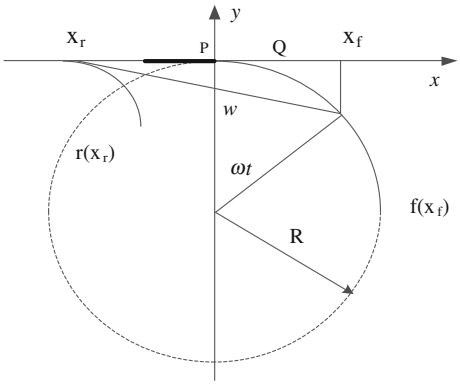


Fig. 4 Circular motion of the long vehicle

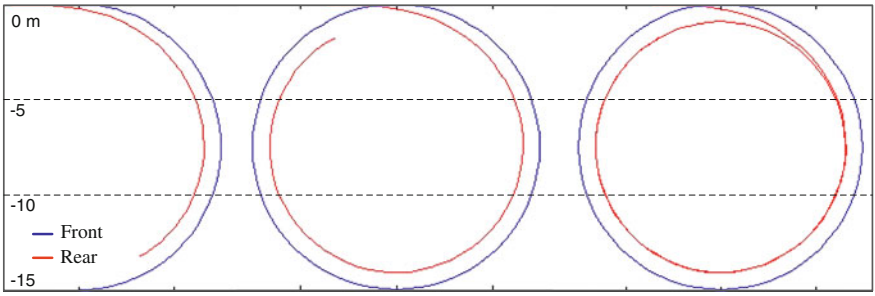


Fig. 5 Motion of Q at different times

circular motion. Notice that from now on, the trajectory of Q is denoted in red while the trajectory of P is in blue. Once Q merges onto the inner circle, it will not deviate from the path as long as P moves along the outer circle.

The parametric function has velocity and time t as its variables, seemingly depending on kinetic parameters. Also, it seems as if this method has a limitation that we have to write the locus curve of P in parametric form, which changes its value over time t. In addition to the shape of the curve of P, are these motion parameters such as velocity, angular velocity necessary to be specified before we calculate the trajectory of Q? The answer is no. Although the Eq. (7) are written in differential form, however, the independent variable here does not have to be time t. During the derivation, the purpose of using differentiation is to obtain the tangent at a specific point on the curve, which is in parametric form and the parameter does not necessarily denote time. Furthermore, the motion of P is not required to move at constant velocity. Consequently, it is very convenient and efficient for trajectory calculation. For this reason, what we need is just the shape of the curve formed by P. So, the parametric equation of P is written using only one parameter later in this article.

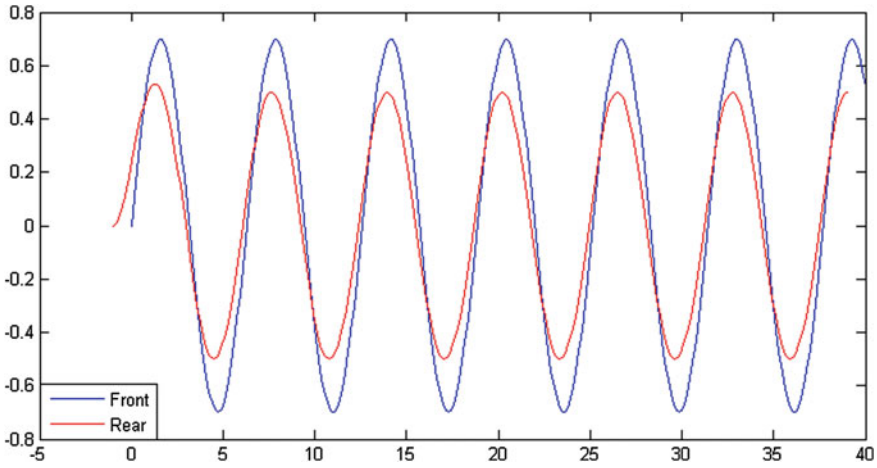


Fig. 6 An example of sinusoidal trajectory

To show that this method can be extended to other types of curves, not just for circular motion, here is another example. As shown in Fig. 6, the motion of P is described as a sinusoidal function (with $L = 1$, $A_f = 0.7$):

$$\begin{aligned} x_f(t) &= t \\ y_f(t) &= A_f \sin(t) \\ x_r(0) &= -1,\ y_r(0) = 0 \end{aligned} \tag{9}$$

With Matlab's ODE45, we can solve it numerically. Figure 6 shows that the trajectory of Q also has a sinusoidal-like shape. Actually, if the trajectory of P is given in an arbitrary shape expressed as discrete coordinates, the analytical expression can be obtained by piecewise fitting methods. For example, piecewise parametric cubic polynomial approximations can be generated automatically for computing from the sampled trajectory data (Plass and Stone, 1983).

2.3 Trajectory of Articulated Vehicles

Articulated vehicles generally include many kinds of vehicles, such as heavy equipment, buses, and even trams and trains. Here, we refer to those vehicles that do not need to run on tracks. Generally, the trailer has one or more axles, while two types of trailer axle placement are popular: single axle and double axle placement, and in most cases, the double axle setup can be treated as equivalent to the single axle mathematically during the calculation of trajectory. For the articulated vehicle shown in Fig. 7, there are more than one trailers towed by a tractor unit, which has a smaller size compared with the trailers. For the reason that it still

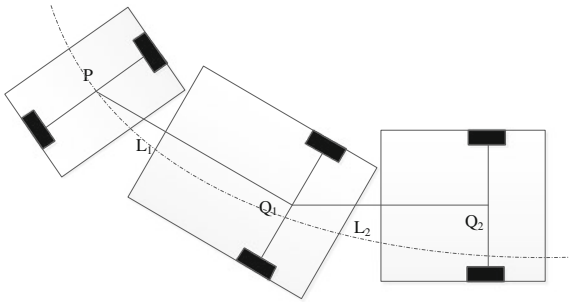


Fig. 7 The scheme of the articulated vehicle

follow the principles described before, the mathematical model still holds true for articulated vehicles. PQ_1, Q_1Q_2, etc., are always perpendicular to the direction of the rear axles of the trailers respectively. Each section, for example, PQ_1, can be treated as a mathematical model described in Fig. 3. In this way, the trajectory of each midpoint of the axle can be calculated one by one by using the method above repeatedly.

2.4 Other Trajectory Analysis Methods

There are various kinds of commercial programs for swept path analysis, and most of them are required to work with AutoCAD for two-dimensional simulation. The swept path analysis software usually works as a plugin for AutoCAD, and allows the customer to draw a user-defined path, and with the preloaded vehicle models in its library and its path calculation algorithm applied (though some simulation algorithms are heuristic, e.g., AutoTURN), the vehicle swept path can be generated accordingly. This way it is capable to calculate and analyze the movement and path of different parts of a vehicle when a turning maneuver happens. The results include the path taken by each wheel and the space needed by the vehicle body during the turn.

Here, we take TURN.LSP project as an example, which is licensed as free software under the terms of the GNU GPL. The basic theory of the program is that a vehicle wheel pair (front and rear) is assumed to be traveling in a circular motion for each computing step. The paths of front and rear wheels are two concentric circles, with the tangent line of the inner circle formed by rear wheel pointing to the front wheel, as Fig. 8 shows.

The initial locations of the front and rear wheels given (P_0 and Q_0), and the final location of the front wheel is known (P_1). The turning angle of rear wheel keeps constant in a circular motion for each step (see Fig. 9).

According to the geometric constrains, the following equations can be derived.

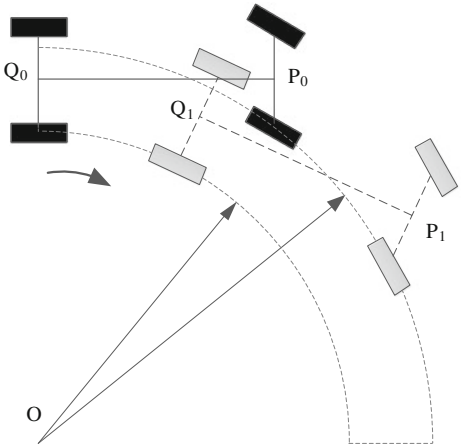


Fig. 8 Relationship between a vehicle wheel pair in each computing step

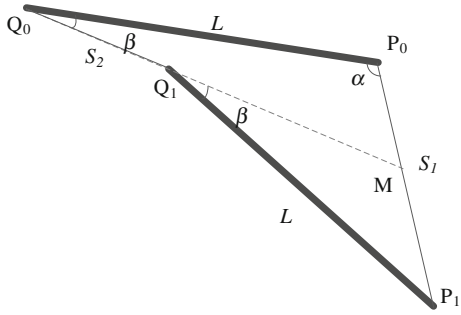


Fig. 9 Mathematical model of vehicle turning in each computing step

$$S_1 \sin(\alpha + \beta) = 2L \sin\beta$$
$$\beta = \tan^{-1}\left(\frac{\sin\alpha}{\frac{2L}{S_1} - \cos\alpha}\right) \tag{10}$$

where S_1, S_2 are the distances between P_0P_1 and Q_0Q_1, respectively, and L is the wheelbase, α and β are the angles as shown in Fig. 9.

From these formulas, obviously point M is the midpoint of P_0P_1, the line between the initial and final location of the front wheel. So, an efficient implementation of this scheme leads to the prediction algorithm, which is also adopted by AutoTrack. To obtain the new position of the rear wheel, first of all, the midpoint of the line P_0P_1 is calculated. Then draw a line connecting this point to the old position of rear wheel, and the new position of the rear wheel should be located along this line. The exact position of it is the intersection of this line with the circle formed by the rear wheel. Its position can be easily determined by calculating the turned angle β. To compare the results obtained from this prediction algorithm, we applied the same parameters and constrains to TURN.LSP.

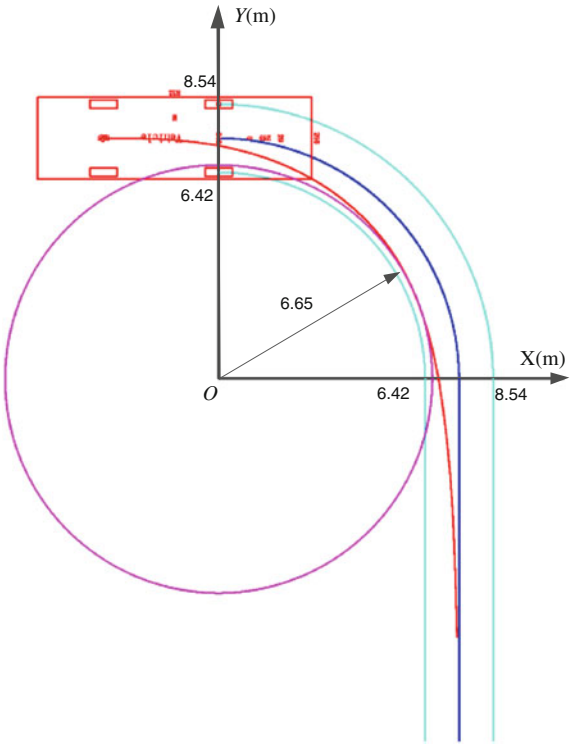


Fig. 10 The trajectory of rear midpoint using TURN.LSP

In AutoCAD we created a block to represent the long vehicle, and default settings were used for computing. Finally the path of the midpoint of the rear axle was generated, as shown in Fig. 10. It is clear that a segment of this curve has the same osculating circle at different positions, which indicates the inner circle formed by the rear midpoint. Here the radius of the inner circle is 6.57, which is different from the prior result due to different algorithms applied.

Because of the assumptions of circular motion and equal turning angle for each step, it is an approximate approach and only deals with constant velocity problem, thus does not reflect the actual situation. However, it is efficient for computation and effective for simulation and can provide visualization for trajectory planning with no strict requirements.

2.5 Reversing of Long Vehicles

In some cases, the vehicle needs to be driven in the reverse direction. Reversing a vehicle serves purposes of changing lanes in traffic, parking, etc. So we require the rear wheels to follow a predefined path instead of steering the vehicle with its minimum turning radius. Now the question becomes how to determine the

trajectories of front wheels and other parts of the vehicle with the trajectories of rear wheels given. With no slippage and other similar assumptions, it also follows the three principles described in Sect. 2.

$$\begin{aligned} \frac{y_{\mathrm{r}}^{'}}{x_{\mathrm{r}}^{'}} &= \frac{y_{\mathrm{f}} - y_{\mathrm{r}}}{x_{\mathrm{f}} - x_{\mathrm{r}}} \\ L^2 &= (x_{\mathrm{f}} - x_{\mathrm{r}})^2 + (y_{\mathrm{f}} - y_{\mathrm{r}})^2 \end{aligned} \tag{11}$$

By solving these two equations with two unknowns, x_{f} and y_{f}, the coordinates of front midpoint can be determined. In this case it does not need any initial conditions because there is a one-to-one correspondence between the coordinates of rear midpoints and those of front midpoints with the direction of the movement is supposed. Once the trajectories of rear midpoint determined, other related trajectories of "feature points" can also be obtained using the same method as the calculating the trajectories of the crane.

3 Application: Mobile Crane

As an application of trajectory planning and simulation for long vehicles, the trajectory of a two-axle mobile crane is analyzed when it is turning around a corner with 90°. The specific model adopted here is LTM1040-2.1, which is made by Liebherr and has all-terrain capability, shown in Fig. 11.

In order to determine whether the mobile crane is allowed to pass through this area and whether the collision will happen throughout the turning process, we set

Fig. 11 Liebherr mobile crane LTM1040-2.1

the front midpoint P moving along the smallest circular turn that the mobile crane is capable of making, i.e., the turning circle. According to the technical data sheet, the turning circle of LTM1040-2.1 has a radius $R = 7.48$ m, with its body length $L = 3.58$ m.

Here, the initial conditions are $x_r(0) = -L$ and $y_r(0) = 0$. Considering that the path of P consists of two parts: circular motion and linear motion, the simulation can be done in two steps, as Fig. 12 shows. The final trajectory is obtained by combining two parts. The radius of the inner circle can be easily determined by checking the data acquired in the second step. Then the radius equals 6.57 m, which is consistent with the geometric relationships.

It is not enough to determine the practical collision situations given only the trajectories of two midpoints of the axles. Ideally, if we know the trajectory profile or envelope of trajectories of the outer outline of the vehicle body when it moves along the predefined path, it is very easy and intuitive to get detailed information and make corresponding decisions. We can choose several "feature points" along the outline, as Fig. 13 shows. These points are chosen because they have a determinant influence on formation of the trajectory envelope. By calculating the trajectories of these points, we can draw a trajectory profile, which can be used to show if the collision will happen by check whether there are intersections between the trajectory profile and the surrounding objects.

Here, the sketch of the mobile crane bottom view just serves as an example to explain how it works, thus not necessarily accurate. We choose two positions at the boom head A, B, and other two positions representing the corners of rear end C, D. Also, more points can be chosen for further precise calculation and verification. The trajectories of these points can be calculated once a series of positions of Q are known. Dimensions marked in Fig. 13 is the geometric relationships between these points and rear midpoint Q. Figure 14 gives us a visualized description of trajectories formed by these chosen points throughout the steering process. According to the diagram, trajectories of point A and D determine the collision situation, and these two lines form an area representing the path swept by the

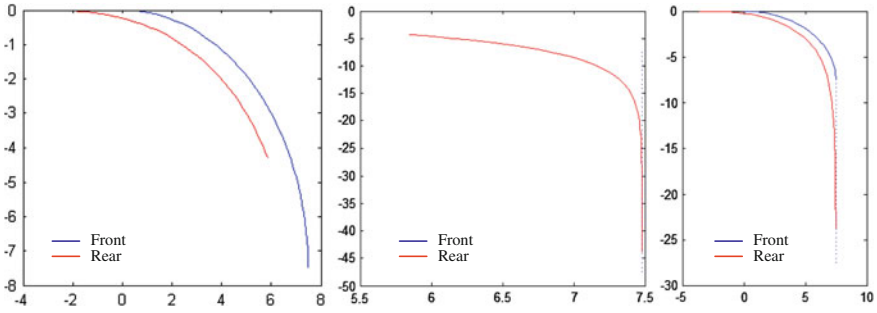


Fig. 12 Numerical simulation of the trajectory

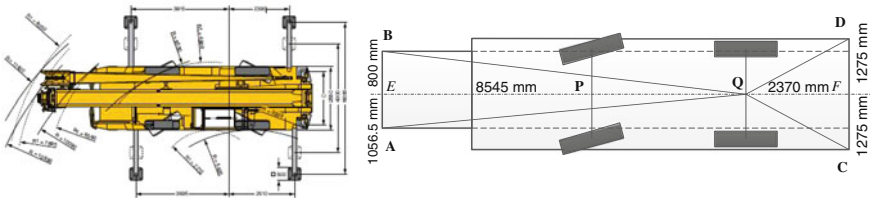


Fig. 13 Several important dimensions of the mobile crane

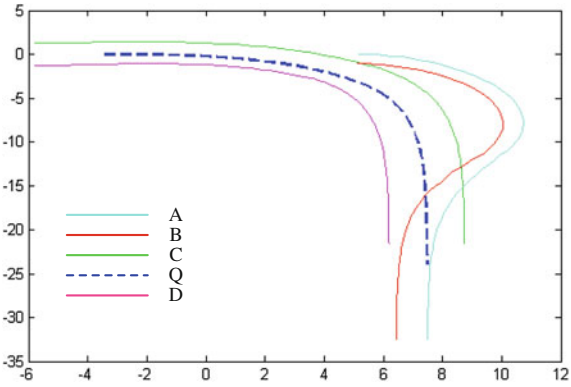


Fig. 14 Trajectories of several feature points

vehicle body. If the surrounding objects, such as barriers, walls, curbs, etc., have intersections with the area between these two lines, the crane will have difficulties passing through the area safely.

4 3D Simulation

For a vehicle with front wheel steering system, the turning sequence around a corner with 90° can be described in Fig. 15.

When the vehicle approaches a corner, it will steer near the outer edge of the incoming road in order to take advantage of the road width. For a long vehicle, the tail out-swing could be significant and this clearance should be consider to prevent collision at the tail of the vehicle. This clearance can be calculated from the vehicle geometry against its wheel position at minimum turning radius.

The position where vehicle should start to steer depends on the front clearance of the vehicle with the outer edge of the outgoing road. There should be a clearance to avoid any collision at the front edge of the vehicle with the surrounding objects. At this point, check should be made whether the inner wheel of the vehicle is able to clear the inner road corner, and this factor is dependent on the width of the outgoing road. With the geometric data of the vehicle and the terrain, the position where driver should start to steer can be calculated.

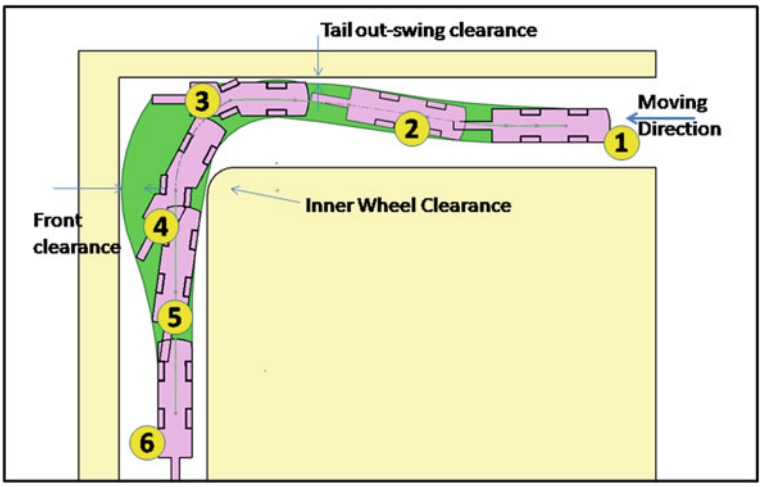


Fig. 15 The turning sequence around a corner with 90°

After the corner, the front steering is release to follow the direction of the outgoing road. The rear wheels will follow the path as discussed earlier in Sect. 2.

The swept area required for the vehicle to make through a corner can be determined from the following:

(1) The vehicle geometry (size, shape, axle centers);
(2) The minimum turning radius of the vehicle (i.e., the maximum steering angle of front wheels);
(3) The rear wheel path of the vehicle, which can be calculated for front wheel steering vehicle. For all wheel steering vehicles, the wheel path solely depends on the driver input and the control system of the vehicle.

Combining this swept area with the road terrain and optimizing with the required clearances as discussed earlier, the possibility of collision can be determined. If there is no overlap, the vehicle is able to turn around the corner. Any overlaps will mean that the vehicle is not able to make through the corner with the given surroundings.

For the mobile crane model adopted in the study, the swept area can be determined by using geometric parameters. Figure 16 illustrates the swept area by the mobile crane under two situations: front wheel steering and all wheel steering.

With the mathematical model described, the problem can be simulated using 3D CAD software with the following input:

(1) The geometry of the mobile crane,
(2) The terrain where the vehicle to be operated, and
(3) The desired path determined with the mathematical mode.

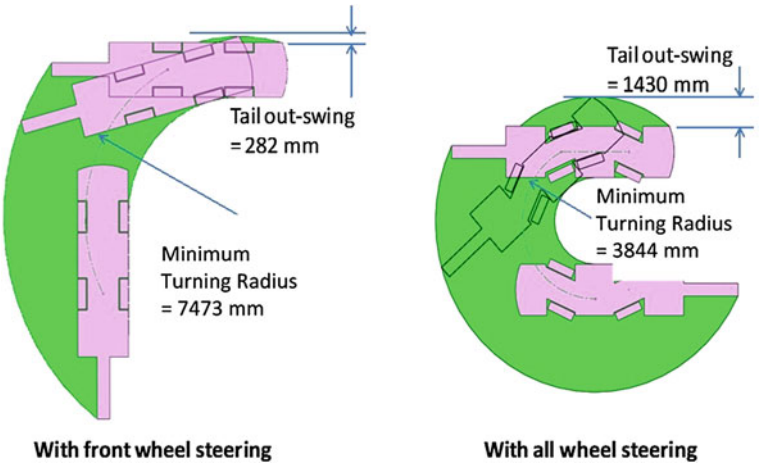


Fig. 16 Maximum displacement of the tail

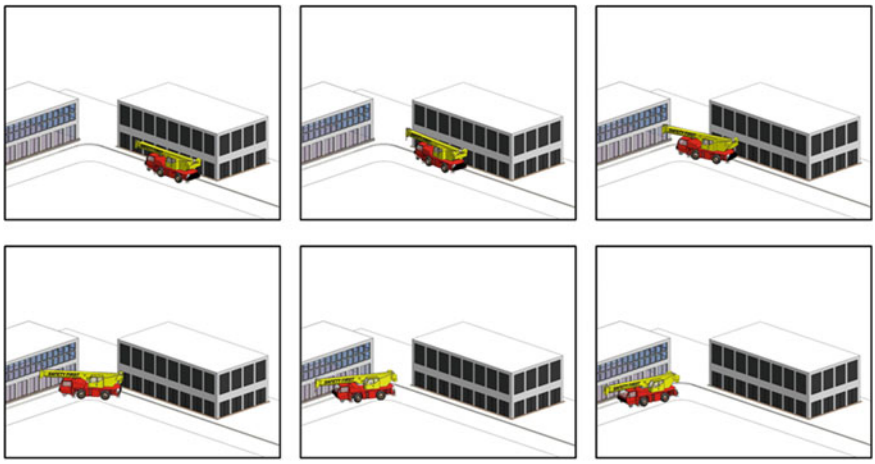


Fig. 17 Simulation of the mobile crane turning around a corner

Two examples of simulation are illustrated with the following figures. Figure 17 is the 3D simulation of the steering process around a corner with 90°. Here there are two buildings serving as the objects. However, in a real world situation, the operator probably encounters more complex environments, scattered with protective edges, barriers, peripheral walls, other vehicles, etc. Figure 18 illustrates the process of maneuvering the mobile crane out of the parking lot. Also, this simulation procedure is reversible and can be used to indicate the whole process of parking the crane.

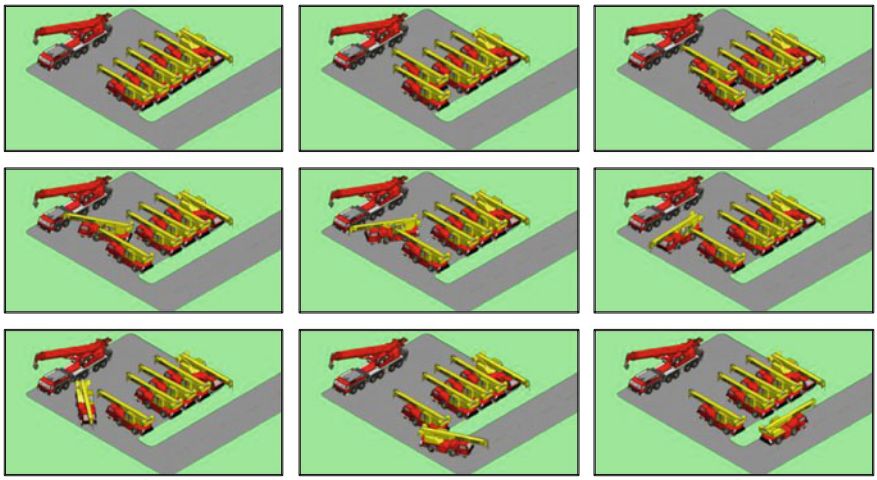


Fig. 18 Maneuvering of mobile crane in parking lot

5 Conclusion

In this chapter, through investigating different mathematical methods for vehicle trajectory calculation, we develop a systematic scheme for modeling long vehicle turning and trajectory calculation using geometrical parameters. Based on the data acquired, 2D numerical analysis and 3D simulation of long vehicles in different environments are conducted to provide accurate and intuitive information for trajectory planning. Also, by comparing with exist swept path analysis method, it shows a better accuracy and agree with the theoretical results, though less time efficient. Considering more realistic factors such as the slippage of the wheel and the road-tire contact, further studies will be carried out to provide more reliable, interactive simulation, as well as optimizing the computations.

References

Baylis J (1973) The mathematics of a driving hazard. Math Gaz 57:22–26

Bender EA (1979) A driving hazard revisited. SIAM Rev 21:136–138

Battelle Team (1995) Comprehen-sive truck size and weight (TS & W) study. Phase 1—synthesis. Roadway geometry and truck size and weight regulations. Working paper 5. Vehicle characteristics affecting safety and truck size. Ohio, Columbus

Bender EA (2000). An introduction to mathematical modeling. Dover Publications, New York

Craig Coulter R (1992) Implementation of the pure pursuit path tracking algorithm. Technical Report CMU-RI-TR-92–01. The Robotics Institute, Carnegie Mellon University, Pittsburg, USA

Dunbar SR, Bosman RJC, Nooij SEM (2001) The track of a bicycle back tire. Math Mag 74(4):273. doi:10.2307/2691097

Erkert TW, Sessions J, Layton RD (1989) A method for determining offtracking of multiple unit vehicle combinations. J For Eng 1(1):9–16

Fancher P, Balderas L (1987) Development of microcomputer models of truck braking and handling. Final Report UMTRI-87–37. The University of Michigan Transportation Research Institute, University of Michigan, USA

Fioretti L, Piacentini F, Liu L, Setekera R, Wang X (2008) Team three optimization of roundabouts. Report for 22nd ECMI modelling week. Netherlands

Freedman HI, Riemenschneider SD (1983) Determining the path of the rear wheels of a bus. Soc Ind Appl Math Rev 25(4):561–567

Hellström T, Ringdahl O (2005) Autonomous path tracking using recorded orientation and steering commands. In: Proceedings of towards autonomous robotic systems 81–87. doi:10.1.1.156.2510

Jesse Lee Farmer (2008) Kinematic analysis of a two-body articulated robotic vehicle. Virginia Polytechnic Institute and State University, Blacksburg

Plass M, Stone M (1983) Curve-fitting with piecewise parametric cubics. Comput Graph 17(3):229–239

Prince GE, Dubois SP (2009) Mathematical models for motion of the rear ends of vehicles. Math Comput Model 49(9–10):2049–2060

Sayers MW (1986) Vehicle offtracking models. Transp Res Rec 1052:53–62

Sweatman P, George R, Tso Y, Ramsay E (1991) A study of heavy vehicle swept path performance. Australian Road Research Board Special Report No, Australia 48

Reliable and Fast Conservative Advancement for Physically Realistic Rigid Body Simulation

XY Zhang

Abstract Conservative Advancement is an efficient technique for performing interactive continuous collision detection in computer graphics, computer game, computer animation, robotics, etc. Conservative advancement has been successfully applied to physically based simulation, robot motion planning for convex, nonconvex, articulated, and deformable objects. In this chapter, some critical technical issues are discussed for implementing reliable and fast conservative advancement. These techniques significantly improve the reliability, physical realisticity, and performance of simulation when conservative advancement is applied to impulse-based dynamics system for rigid bodies. These techniques are easy to implement and have been integrated into a CA-based, open source project.

Keywords Collision detection · Conservative advancement · Rigid body

1 Introduction

At a broad level, collision detection can be categorized into two types, *discrete collision detection* and *continuous collision detection*. Discrete collision detection algorithms check for interferences (interpenetration or not) at fixed time instances. When objects move rapidly or objects are thin, discrete collision detection algorithm may miss collisions for some time intervals (see Fig. 1a for an example). Even though collisions can be detected, this method merely reports interpenetration or collision-free. In order to handle responses in dynamics simulation of rigid bodies, two techniques are typically used in discrete collision detection: penetration depth or backtracking method. However, computing penetration depth is

XY. Zhang (✉)
Nanjing University, Nanjing, China
e-mail: zhangxy@cs.unc.edu

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_7,

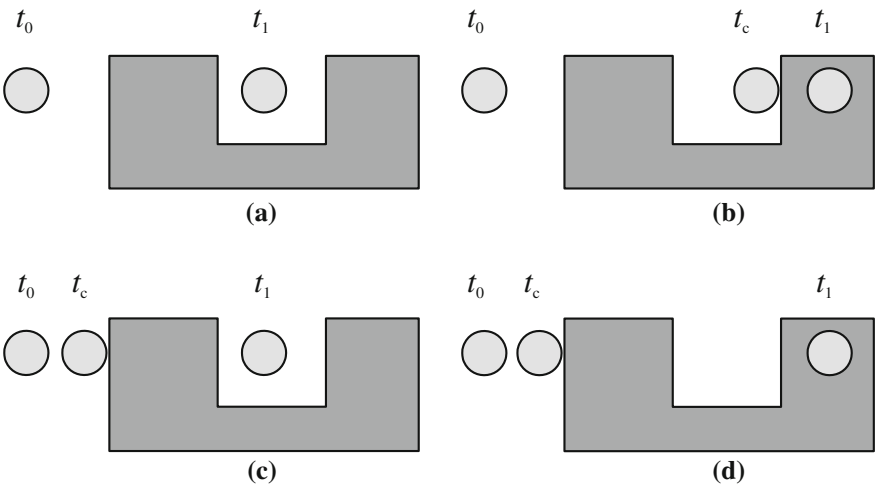


Fig. 1 **a** Discrete collision detection algorithms may miss collisions. **b** Backtrack method may incorrectly report the time of contact. **c** Continuous collision detection is able to avoid the problem occurring in (**a**). **d** Continuous collision detection is able to avoid the problem occurring in (**b**) because it finds the first time of contact

difficult and expensive (Je et al. 2012). Even though penetration depth can be determined, it is often quite difficult to know the contact position that is very important for computing the responses of dynamics simulation for rigid bodies in order to avoid perceptual artifacts. Backtracking method can approximate the first time of contact with a bisection scheme in which discrete collision detection algorithms are repeatedly invoked. It avoids expensive penetration computation, but it can still fail in some every simple case such as the one shown in Fig. 1b. Moreover, the computational cost can be still very high for complex models. For the difficulties of the use of discrete collision detection in rigid body dynamics, we refer readers to (Redon et al. 2004; Redon 2004, 2009)

As illustrated in Fig. 1c and d, using continuous collision detection, however, expensive penetration and backtracking can be avoided, and moreover the contact witness features can be obtained after moving the bodies to the time of contact. This can facilitate the dynamics simulation of rigid bodies.

In our recent work (Zhang et al. 2006), we presented a continuous collision detection algorithm based on conservative advancement (CA) for rigid polyhedra. We discussed some applications using the algorithm. In particular, we applied our CA-based continuous collision detection to dynamics simulation of rigid bodies. However, in order to generate physically realistic dynamics simulation for rigid bodies, some technical details of the CA-based continuous collision detection are needed to clarify further. Here, we elaborate on the algorithm such that the algorithm can be effortlessly integrated into existing dynamics simulation of rigid bodies.

The rest of this chapter is organized as follows. In Sect. 2, we briefly describe the recent work on CA-based continuous collision detection. In Sect. 3, we explain

how the CA-based continuous collision detection is used for the dynamic simulation of rigid bodies. In Sect. 4, we show some experimental results of our algorithm in different benchmarking scenarios and conclude this chapter in Sect. 5.

2 Overview of CA-Based CCD

Conservative advancement is a decent and simple method for approximating the first time of contact between objects. The original literatures discuss the use of conservative advancement only for convex polyhedra (Mirtich 1996); however, our recent work extended conservative advancement to nonconvex polyhedra (Zhang et al. 2006) and even for articulated bodies consisting of nonconvex polyhedra (Zhang et al. 2007). The simplicity of conservative advancement lies in the fact that it uses closest distance computation for querying a lower bound of the time of contact for continuous collision detection.

Let d(A, B) denote a closest distance between two bodies A and B. The time of contact is the minimum value of t that satisfies $t \in [0, 1]$ and $d(A(t), B(t)) = 0$. Thus, the problem of CA-based continuous collision detection can be formulated as follows:

$$\{\min(t) | t \epsilon [0, 1], d(A(t), B(t) = 0)\} \tag{1}$$

Conceptually, computing the time of contact is a root-finding problem (see Fig. 2).

We can simply classify the root-finding algorithms into two categories: open methods and bisection methods. Open methods are based on the formulas that require a single starting point. The iterative formulas are repeatedly evaluated until a root is reached. The work based on conservative advancement (Mirtich 1996, 2000; Zhang et al. 2006, 2007) belongs to this category. Bisection methods are based on the brackets that contain the root. If the root is found in a bracket, this bracket is subdivided into two sub-brackets and then the root is recursively looked

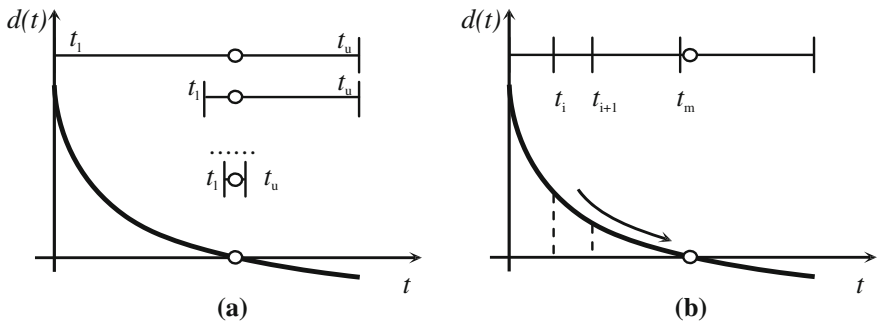


Fig. 2 Root-finding of equations. *Left* bisection methods; *Right* open methods

for in these sub-brackets. The root-finding process stops as soon as the bracket which contains the root is small enough. The work-based interval arithmetic (Redon et al. 2002) belongs to this category.

Open methods often converge more quickly than bisection methods. For example, for a translational motion, conservative advancement requires only one iterative step to find the time of contact

2.1 Conservative Advancement for Convex Polytopes

Analog of open methods, conservative advancement of convex polytopes uses Eq. 2 to compute a lower bound of the time of contact t_c, while repeatedly advancing objects A and B by Δt. More specifically, as shown in Fig. 3, let **n** be the closest direction vector that realizes d(A(t), B). Then, object A can advance from time t to time $t + \Delta t$ without collision.

In addition, if $\Delta t > 1$, $t_c > 1$ and the pair of relevant bodies do not collide. If $\Delta t < 1$ and $t_c < 1$ at some iteration, the relevant bodies move by the time step Δt to the new position. Then a new time step is computed and the new t_c is a better lower bound. This process repeats until the relevant bodies are close enough to each other. We may say that the pair of relevant bodies is in contact at t_c

$$\Delta t = \frac{\mathrm{d}(\mathrm{A}(t), \mathrm{B}(t))}{\bar{u}} \qquad t_c = \sum \Delta t_i \tag{2}$$

where $\bar{u}$ is an upper bound relative velocity along the closest direction (**n**) of two bodies. Note that the smaller value of $\bar{u}$, the larger CA step size Δt we can have. Thus, we use the following formula for computing a tight upper bound of $\bar{u}$.

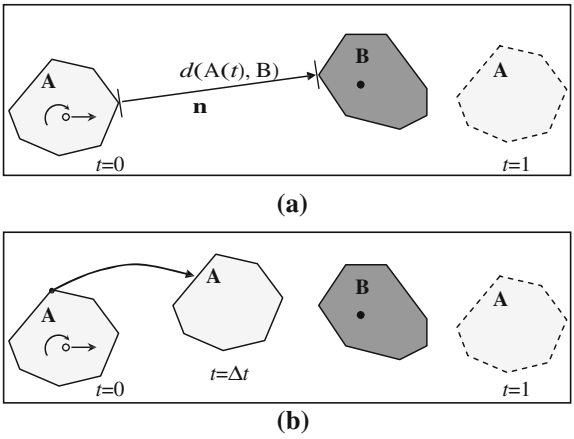


Fig. 3 Conservative advancement for convex polytopes

$$\bar{u} = \left|\mathbf{v}^A \cdot \mathbf{n}\right| + \left|\mathbf{n} \times \boldsymbol{\omega}^A\right| \max_i \left(r_i^A\right) \tag{3}$$

Without loss of generality, we assume that A is movable and B is fixed in space since we can find relative rigid transformation with respect to each other.

A linear motion is used for determining the constant motion parameters **v** and $\boldsymbol{\omega}$. More specifically, given two discrete configurations $\mathbf{q}_0 = (\mathbf{R}_0, \mathbf{T}_0)$, $\mathbf{q}_1 = (\mathbf{R}_1, \mathbf{T}_1)$ with translational (**T**) and rotational (**R**) components, the linear motion **M**(t) is represented with respect to time t as:

$$\mathbf{M}(t) = \begin{pmatrix} \mathbf{R}(t) & \mathbf{T}(t) \\ (0,0,0) & 1 \end{pmatrix} \tag{4}$$

Then

$$\mathbf{T}(t) = \mathbf{T}_0 + t\mathbf{v}$$

$$\mathbf{R}(t) = \cos(|\omega|t) \cdot \mathbf{A} + \sin(|\omega|t) \cdot \mathbf{B} + \mathbf{C}$$

Here, $\mathbf{v} = \mathbf{T}_1 - \mathbf{T}_0$ is the constant translational velocity of the center of mass (COM). ω is angular velocity evaluated by the relative orientation of $\mathbf{R}_0$ and $\mathbf{R}_1$. Furthermore,

$$\mathbf{A} = \mathbf{R}_0 - \omega\hat{\omega}^{\mathrm{T}} \cdot \mathbf{R}_0$$

$$\mathbf{B} = \hat{\omega}^* \cdot \mathbf{R}_0$$

$$\mathbf{C} = \hat{\omega}\hat{\omega}^{\mathrm{T}} \cdot \mathbf{R}_0$$

where $\hat{\omega}^*$ is the skew symmetric matrix; i.e., if $\hat{\omega} = (\hat{\omega}^x, \hat{\omega}^y, \hat{\omega}^z)^{\mathrm{T}}$

$$\hat{\omega}^* = \begin{pmatrix} 0 & -\hat{\omega}^z & \hat{\omega}^y \\ \hat{\omega}^z & 0 & -\hat{\omega}^x \\ -\hat{\omega}^y & \hat{\omega}^x & 0 \end{pmatrix}$$

2.2 CA for Nonconvex Polyhedra

The most important concept of conservative advancement is distance computation. The distance between two convex polytopes can be efficiently computed either using LC (Lin 1993) or GJK (Gilbert et al. 1988) algorithm. Moreover, the closest direction returned by distance computation algorithms can be further used to calculate the tight upper bound of relative velocity between two bodies. This technique can be generalized to nonconvex objects using convex hull hierarchies or convex hull trees (Ehmann and Lin 2001). Using convex hull hierarchies, the local distances between two nonconvex bodies can be calculated efficiently and the pair of nodes in convex hull trees that realize every local distance is remembered

and called front nodes. Note that every node of a convex hull tree is convex. The union of these front nodes is a super set of the union of convex pieces thanks to the convex hull property of convex hull tree. For nonconvex bodies, conservative advancement is only performed to these front nodes. Because any pair of front nodes from two bodies are convex, the upper bound of relative velocity can be very tight by taking the closest direction **n** and projecting its translational and rotational velocities onto **n** (Eq. 3). Overall, the algorithm works like the one in Algorithm. 1.

Distance. In Algorithm 1, this is a procedure for distance computation between two general polyhedra. Returned values are a set of front nodes $\left(h_i^A, h_i^B, d_i, n_i\right)$. A local distance and its closest direction (d_i, n_i) is implemented by the pair of convex nodes $\left(h_i^A, h_i^B\right)$. As mentioned early, a well-known BVH-based technique (Ehmann and Lin 2001) is used to implement the procedure. As preprocess, the algorithm decomposes a given, general polyhedron into convex pieces and recursively builds a convex hull tree where each node in the tree corresponds to a convex hull of its children nodes. Note that the convex decomposition scheme employed in (Ehmann and Lin 2001) is merely surface decomposition in which the union of convex pieces covers only the boundary of the object, and this is sufficient for distance calculation. At run-time, starting from the root nodes of two convex hull trees, the algorithm simultaneously traverses the two trees while performing Voronoi marching on the tree nodes and calculating their distance as well as their associated closest features. The recursive traversal continues until the closest distance cannot be further reduced.

UpperBound. In Algorithm 1, this is a procedure for the upper bound computation for a front node. Formula (3) (Sect. 2.1) is applied to efficiently compute such a bound.

Algorithm 1 CA$(A, B, \mathbf{q}_0, \mathbf{q}_1)$

Compute an interpolating motion $\mathbf{M}(t)$ from $\mathbf{q}_0, \mathbf{q}_1$

$t_c \leftarrow 0, I \leftarrow 0$
repeat
  $(h_i^A, h_i^B, d_i, n_i) \leftarrow$ Distance$(\mathbf{M}(t_c)A, \mathbf{M}(t_c)B)$
  if $\min(d_i) < \varepsilon$ **then**
    break
  end if

  for each pair of front nodes (h_i^A, h_i^B) **do**
    $\bar{u} \leftarrow$ UpperBound(h_i^A, h_i^B, n_i)
    $\Delta t_i \leftarrow d_i / \bar{u}$
  end for

  $t_c \leftarrow t_c + \min(\Delta t_i)$
  $I \leftarrow I + 1$
until $t_c < 1$ and $I < I_{CA}$

return $\min(t_c, 1)$

3 CA-Based CCD for Rigid Body Dynamics

One of the applications of continuous collision detection is rigid body dynamics simulation. Since, most early collision detection algorithms are discrete, they only report interpenetration or collision-free. Even though discrete collision detection together with penetration depth computation or backtracking method is a typical means for computing the responses of rigid body dynamics simulation for many years, continuous collision detection could be a more natural way to do so. Continuous collision detection can avoid expensive penetration depth computation and recursive backtracking, and moreover the contact witness features can be obtained after moving the bodies to the time of contact (Algorithm 1).

During dynamics simulation, once a collision is detected in the current time interval $[t_i, t_{i+1}]$, the bodies are moved to the time of contact t_c where $t_c \in [t_i, t_{i+1}]$, the simulation resumes from t_c and the time interval of collision detection becomes $[t_c, t_{i+2}]$. However, from Algorithm 1, once the bodies move to the time of contact it will be impossible for those bodies which realized the time of contact to advance to the next time instance. The main reason is the position at t_c of the current time interval is the initial position of the next time interval $[t_{i+1}, t_{i+2}]$, where $t_{i+1} = t_c$. The closest distance at the initial position of the time interval $[t_{i+1}, t_{i+2}]$ is less than ε once the bodies are repositioned to t_c, so that it will report the time of contact 0 for the time interval $[t_{i+1}, t_{i+2}]$ based on Algorithm 1. If the relevant bodies do not move, the closest distance at that position will not change. If the closest distance is always less than ε, the time of contact persistently is reported to be 0. The relevant bodies will be stuck there permanently. An example is given in Fig. 4.

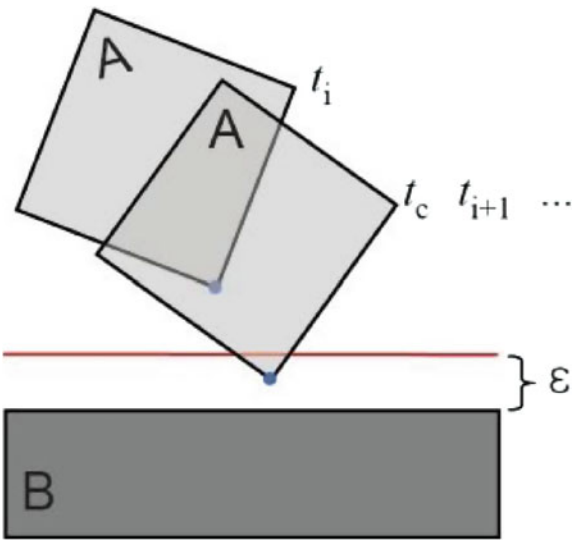


Fig. 4 The body A will be stuck at the time of contact t_c once a collision is detected using Algorithm 8.1

Another problem with Algorithm 1 is that it may introduce numeric errors. To avoid numeric errors, some other degenerate cases (divided by zero, or just contact) need careful treatment.

The main goal of this chapter is to discuss some technical issues and implementation details for designing a practical rigid body dynamics simulator by adapting Algorithm 1. With some adaption of the original algorithm (Algorithm 1), we will demonstrate that CA-based continuous collision detection is applicable to rigid body dynamics simulation.

3.1 Motion Interpolation

Unit quaternion is usually regarded as the best representation for the orientations for rigid body dynamics simulation, so that unlike our previous work where rotation matrices are used to represent orientations, we use unit quaternion in this chapter. Moreover, unlike other applications where only initial and final configurations are known, some motion parameters, including linear and angular velocities are known in prior in dynamics simulation, so that the linear interpolating motion $\mathbf{M}(t)$ can be determined as follows (Mirtich 1996):

$$\mathbf{T}(t) = \mathbf{T}_0 + \mathbf{v}t$$

$$\mathbf{Q}(t) = \mathbf{Q}_0 + \dot{\mathbf{Q}}t$$

where the derivative of a quaternion $\dot{\mathbf{Q}} = \frac{1}{2}\omega\mathbf{Q}_0$.

3.2 CA for Rigid Body Dynamics

Redon (2004) and Redon et al. (2002) suggest using a security distance to prevent interpenetration during rigid body dynamics simulation. In our previous work, we follow this suggestion (see Algorithm 1 and Fig. 4). Here, we use ε_s to denote a security distance. But we suggest using one more distance threshold to indicate if bodies are in contact. Such a distance threshold is denoted as ε_c. When a collision is detected, the distance between contact points is less than $\varepsilon_s + \varepsilon_c$.

As shown in Fig. 5, our goal is to obtain a new ε_c so that the bodies are at collision-free position in the next time interval. We suggest that the contact points are merely allowed falling into the distance interval $[\varepsilon_s + \varepsilon_c/2, \varepsilon_s + \varepsilon_c]$. This guarantees that the bodies are at collision-free positions once ε_c is halved.

With a minor adaption of the formula (2), the second condition can be guaranteed. For each CA step,

$$\Delta t \leftarrow \frac{(d - \varepsilon_s - \varepsilon_c/2)}{(\mu)} \tag{5}$$

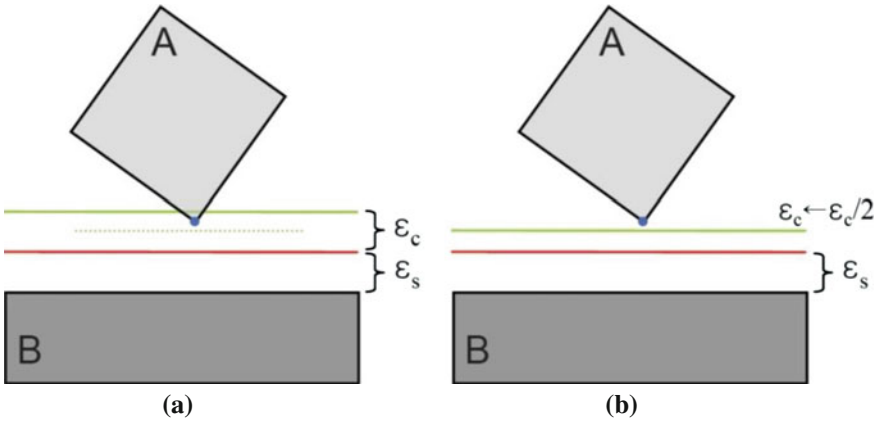


Fig. 5 A security distance (ε_s) is maintained to prevent interpenetration. A contact distance (ε_c) is used to indicate if bodies are in contact or not

This criterion guarantees that a contact distance $d > \varepsilon_s + \varepsilon_c/2$. Meanwhile, CA itself guarantees that a contact distance $d > \varepsilon_s + \varepsilon_c$ at the first time of collision. These two criteria guarantee that a contact point can be exposed to a collision-free position once ε_c is halved.

$$d \in [\varepsilon_s + \varepsilon_c/2, \varepsilon_s + \varepsilon_c] \tag{6}$$

In summary, we use the following three sub-algorithms to query the time of contact for rigid body dynamics simulation.

Table 1 summarizes the notations used in the algorithms.

Algorithm 2 For each body, we first compute its linear motion with constant translational and angular velocities produced by dynamics simulator. Using the resulting linear motion and the technique: Interval Arithmetic (IA) or Taylor Models (TM), dynamics bounding volume culling is performed first. Most of the pairs are culled away during this stage. For those pairs that survive during the dynamic bounding volume culling, conservative advancement is applied (see Algorithm 3). For each pair, its time of contact is compared with the global time of contact (t_c^g) which is initialized to 1.0 at the beginning of the algorithm. If $t_c > t_c^g$, the time of contact cannot be implemented by the current pair; otherwise if $t_c < t_c^g$, t_c^g is updated with t. The final t_c^g is the time of contact for the whole scene. If $t_c^g = 1$ and fs = false, there is no collision. If $t_c^g < 1$, there is a collision at t_c^g. If $t_c^g = 1$ and fs = true, there is a collision at $t = 1.0$. Here, we use "fs" to record if the bodies are in contact at t_c^g.

Once a collision is detected, we halve the contact distance $\varepsilon_c \leftarrow \varepsilon_c/2$. This enables the simulator to evolve to the next time instance. However, if no collision is detected, we increase ε_c by doubling it.

Table 1 A summary of the notation used in the algorithms

h	Convex hull or convex node
u	Linear velocity
v	Linear velocity of center of mass
ω	Angular velocity
r	Rotational radius with respect to center of mass
$\mathbf{M}(t)$	Interpolating motion
$\mathbf{T}(t)$	Interpolating position
$\mathbf{Q}(t)$	Interpolating orientation
t_c^g	The first time of collision of all pairs of objects
t_c	The first time of collision of a pairs of objects
Δt	The CA time step of a pairs of objects
Δt_i	The CA time step of a pairs of convex nodes
n	Contact normal
d	Contact distance
μ	Up bound of Interpolating motion
ε_{s}	Absolute security distance
ε_{c}	Absolute contact in/out distance
ε_s^r	Relative security distance
ε_c^r	Relative contact in/out distance
$\overline{\varepsilon_c^r}$	Default maximum Relative contact in/out distance
$\underline{\varepsilon_c^r}$	Default minimum Relative contact in/out distance
$\overline{\mathrm{fs}}$	Falling into safe distance
PCPs	Potentially colliding pairs
I	Number of iterations
ICA	Maximum number of iterations of CA procedure
ICHCA	Maximum number of iterations of CHCA procedure

Algorithm 3 For each pair that survives in bounding volume culling, we perform ordinary CA using this algorithm. However, prior to performing CA iterations using convex hull hierarchies, we execute CA between root convex hulls of those bodies (see Algorithm 4). The resulting time of contact of Algorithm 4 will be the starting point for the ordinary CA using convex hull hierarchies. The ordinary CA is repeatedly executes until the distance is less than $\varepsilon_s + \varepsilon_c$ (a collision is detected) or $t_c < t_c^g$ (a collision cannot be realized by this pair).

Algorithm 4 Before performing ordinary CA iterations using convex hull hierarchies, we first execute CA between root convex hulls of those bodies survived in dynamics bounding volume pruning. This step only involves the distance computation between a pair of convex hulls, so that the computation is much cheaper than an ordinary CA using convex hull hierarchies. If the first time of collision computed during this step is greater than t_c^g, the exact first time of collision implemented by the pair must be greater than t_c^g and the computation need not proceed. If the first time of collision is less than t_c^g and even it could be 0, such a lower bound of TOC can be used as an initial time instance from which the CA starts for the pair of convex hull hierarchies. Note that it is possible that a pair of

convex hulls can intersect at initial position. Here, the procedure *Distance* is responsible for detecting intersection. If an intersection is found, it reports $d < 0$.

Algorithm 5 This is a routine for calculating a single time step using Eqs. (3) and (5). The time step guarantees that the bodies can advance to a new position without collision. Moreover, the condition $(d > \varepsilon_s + \varepsilon_c/2)$ is guaranteed as well.

Algorithm 2 QueryTOC()

for each body **do**
  Compute an interpolating motion $\mathbf{M}(t)$ from v, ω
end for

Dynamic BV culling (IA or TM) $\rightarrow$ PCPs

$t_c^g \leftarrow 1$, fs$\leftarrow$false
for each pair (A,B) in PCPs **do**
  $\varepsilon_s \leftarrow \varepsilon_s^r \min(r^A, r^B)$
  $\varepsilon_c \leftarrow \varepsilon_c^r \min(r^A, r^B)$
  $(t_c, \text{fs}) \leftarrow \text{CA}(\text{A}, \text{B}, t_c^g, \varepsilon_s, \varepsilon_c)$
  $t_c^g \leftarrow \min(t_c, t_c^g)$
end for

if $t_c^g < 1$ or fs=true **then**
  $\varepsilon_c^r \leftarrow \max(\varepsilon_c^r/2, \underline{\varepsilon_c^r})$
else
  $\varepsilon_c^r \leftarrow \min(2\varepsilon_c^r, \overline{\varepsilon_c^r})$
end if

return colliding features at t_c^g

Algorithm 3 $\text{CA}(A, B, t_c^g, \varepsilon_s, \varepsilon_c)$

$t_c \leftarrow \text{CHCA}(A, B, \varepsilon_s, \varepsilon_c)$

$I \leftarrow 0$, fs $\leftarrow$ false
while $t_c < t_c^g$ and $I < I_{CA}$ **do**
  $(h_i^A, h_i^B, d_i, n_i) \leftarrow \text{Distance}(\mathbf{M}(t_c)A, \mathbf{M}(t_c)B)$
  if $\min(d_i) < \varepsilon_s + \varepsilon_c$ **then**
    fs $\leftarrow$ true
    break
  end if

  for each pair of (h_i^A, h_i^B) **do**
    $\Delta t_i \leftarrow \text{StepCA}(h_i^A, h_i^B, di, ni, \varepsilon_s, \varepsilon_c)$
  end for

  $t_c \leftarrow t_c + \min(\Delta t_i)$
  $I \leftarrow I + 1$
end while

return (t_c, fs)

Algorithm 4 CHCA$(A, B, t_c^g, \varepsilon_s, \varepsilon_c)$

$I \leftarrow 0, t_c \leftarrow 0$
$h^A \leftarrow CH^A, h^B \leftarrow CH^B$
repeat
  $(d, n) \leftarrow$ Distance$(\mathbf{M}(t_c)h^A, \mathbf{M}(t_c)h^B)$
  if $d <= 0$ **then**
    return 0
  else
    if $d < \varepsilon_s + \varepsilon_c$ **then**
      break
    end if
  end if

  $\Delta t \leftarrow$ StepCA$(h^A, h^B, d, n, \varepsilon_s, \varepsilon_c)$

  $t_c \leftarrow t_c + \Delta t$
  $I \leftarrow I + 1$
until $t_c < t_c^g$ and $I < I_{CHCA}$

return t_c

Algorithm 5 StepCA$(h_1, h_2, d, n, \varepsilon_s, \varepsilon_c)$

$\mu \leftarrow$ UpperBound(h_1, h_2, n)
$d' \leftarrow d - \varepsilon_s - \varepsilon_c/2$
if $\mu > 0$ and $d' > 0$ **then**
  return d'/μ
else
  return 0
end if

3.3 Rigid Body Dynamics

There are three main approaches to computing contacts: penalty methods, impulse-based methods, and constraint-based methods. The former two are local and thus simpler and the last one is global and more complex (Guendelman et al. 2003).

Penalty methods cause a stiff spring to the contact points to pull the bodies apart. When the bodies are no longer together, the spring is destroyed. This may be simple to implement but it is hard to choose the spring constants. Usually, spring constants that must be chosen are very high causing the differential equations describing them to become stiff meaning that very small time steps must be used to evaluate them in order to achieve any accuracy whatsoever. In addition, this method causes slight penetrations. This method is local. It does not care about any other points in the whole simulation but the single pair of contact points. *Constraint-based methods* are global. They take into account all contact points at a given instant in time in order to compute the forces required at all contacts. There

are various publications available to the interested reader. Details can be found in a paper by Baraff. *Impulse-Based Methods* are local. All contacts are modeled as collisions. If friction and restitution are incorporated into the collision model, then the resulting impulse trains can generate persistent phenomena such as rolling, sliding, and settling. The advantage of this method over the constraint methods is that the computation is local and very simple. There is no complex system to solve. A drawback is that only a few persistent contacts can be modeled. Static friction is not very well defined in this model. Details can be found in the PhD thesis of Brian Mirtich (1996). In this chapter, we use impulse-based methods because of its simplicity.

3.4 Discussion

An open problem with conservative advancement is the difficulty of handling hovering cases. Assume that a body A with high angular velocity hovers above the body B and the distance is slightly greater than $\varepsilon_s + \varepsilon_c$. The convergence of CA algorithm can be very slow because the time step resulted from the formula is very small.

Then the accumulated time of contact $t_c \in \left[0, t_c^g\right]$, but fs = false. The simulation becomes extremely slow when such a case occurs.

4 Implementation and Experimental Results

We will now explain some implementation details and show experimental results.

4.1 Implementation

We have implemented our CA-based CCD algorithm and rigid body dynamics simulator using C++ on Windows XP, equipped with Intel P4 3.6 GHz CPU and 2 GB main memory. We have used a public-domain proximity library, SWIFT++ for distance calculation and convex decomposition for nonconvex polyhedra, and have modified it to serve our purpose.

4.2 Experimental Results

Rigid Body Dynamics for Bunnies. We have tested our algorithm for rigid body dynamics. As shown in Fig. 6, one bunny drops onto the other bunny under gravity and is bounced away because of impact. Each bunny consists of 26 K triangles. Each simulation step provides $\mathbf{q}_0$, $\mathbf{q}_1$ of the blue moving bunny, and we plug these into our CCD algorithm and measure the performance. Moreover, in order to create artificial interpenetration at $\mathbf{q}_1$ from this simulation (the simulation itself is collision-free all the time), we slightly modify the $\mathbf{q}_1$ before impact, and thus we can invoke our CCD algorithm to find t_c.

Rigid Body Dynamics for Rings (Fig. 7). We conduct similar dynamics experiments for rings consisting of 10 K triangles, each.

Rigid Body Stack (Fig. 8). We integrated our algorithm into 2DBox (Catto 2012) and performed dynamics simulation for stack of rigid bodies.

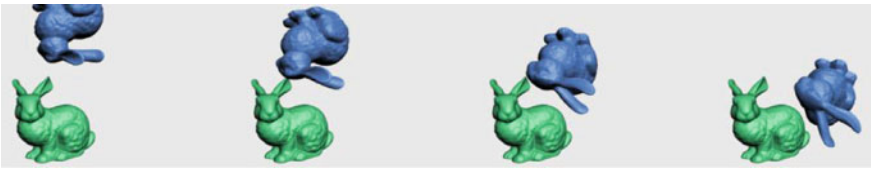


Fig. 6 Rigid body dynamics for bunnies. A *blue* bunny falls onto a *green* one and is bounced off

Fig. 7 Rigid body dynamics for *rings*. A *blue ring* falls onto a *green* one and is bounced off

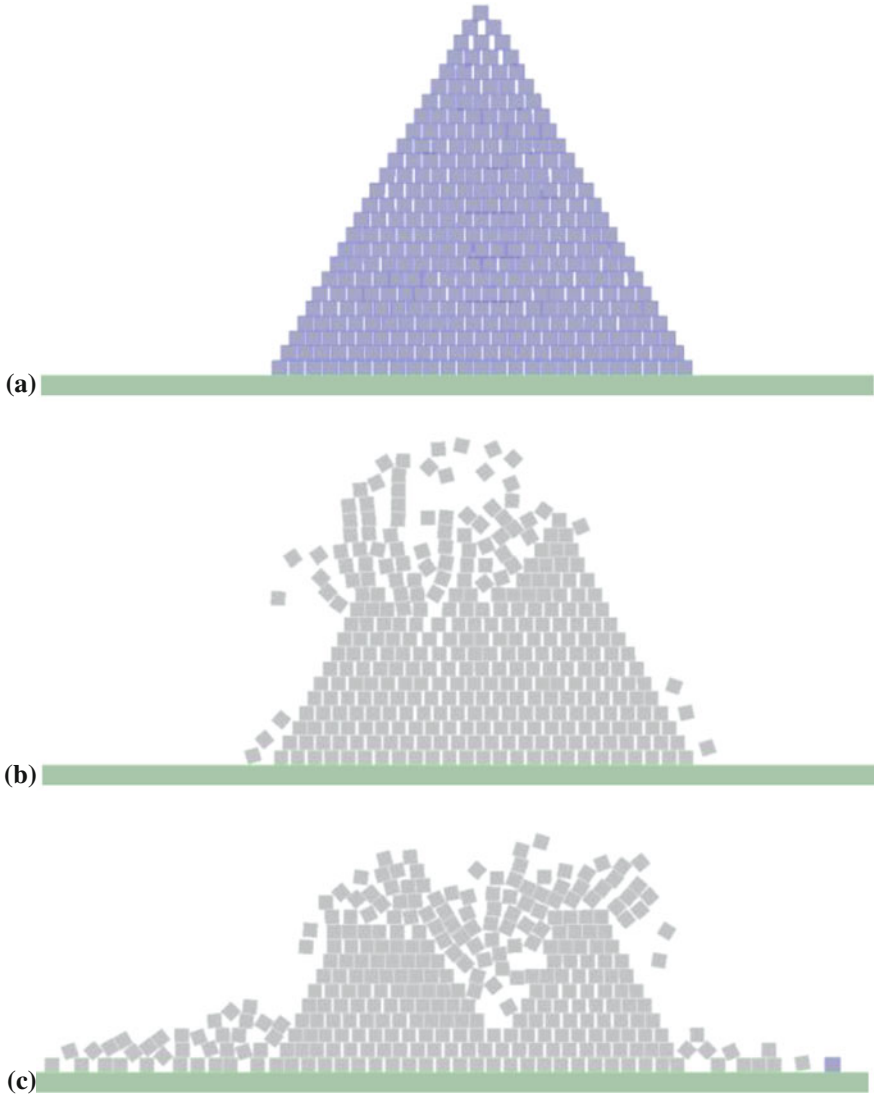


Fig. 8 Rigid body stack

5 Conclusion

We have discussed some technical issues arising in implementing reliable and fast conservative advancement-based continuous collision detection algorithms, which is very important for performing interactive and physically realistic dynamics

simulation for rigid bodies. We have demonstrated that our solutions are able to significantly improve the reliability, physical realisticity, and performance of simulation.

References

Catto E (2012) http://box2D.org
Ehmann S, Lin MC (2001) Accurate and fast proximity queries between polyhedra using convex surface decomposition. Eurographics 20(3):500–510
Gilbert EG, Johnson DW, Keerthi SS (1988) A fast procedure for computing the distance between complex objects. Int J Robot Autom 4(2):193–203
Guendelman E, Bridson R, Fedkiw R (2003) Nonconvex rigid bodies with stacking. ACM Trans Graph (SIGGRAPH) 22(3):871–878
Je C, Tang M, Lee Y, Lee M and Kim YJ (2012) PolyDepth: realtime penetration depth computation using iterative contact-space projection. ACM Trans Graph 31(1):5:1–5:14
Lin MC (1993) Efficient collision detection for animation and robotics. Ph.D. Dissertation, University of California, Berkeley
Mirtich BV (1996) Impulse-based dynamic simulation of rigid body systems. Ph.D. Dissertation, University of California, Berkeley
Mirtich BV (2000) Timewarp rigid body simulation. ACM SIGGRAPH 193–200
Redon S (2009) Continuous collision detection. In: Lin MC, Otaduy M (eds) Haptic rendering: foundations, algorithms and applications. AK Peters, Ltd, London
Redon S, Kheddar A, Coquillart S (2002) Fast continuous collision detection between rigid bodies. Computer Graphics Forum (Proc. of Eurographics) 21(3):279–288
Redon S, Kim YJ, Lin MC, Manocha D (2004) Interactive and continuous collision detection for avatars in virtual environments. IEEE Virtual Reality 117–283
Redon S (2004) Continuous collision detection for rigid and articulated bodies. ACM SIGGRAPH Course
Zhang XY, Lee M, Kim YJ (2006) Interactive continuous collision detection for non-convex polyhedra. Pac Graph 749–760
Zhang XY, Redon S, Lee M, Kim YJ (2007) Continuous collision detection for articulated models using taylor models and temporal culling. ACM Trans Graph (SIGGRAPH) 26(3):15

The Use of Virtual Worlds and Serious Gaming in Education

Wim Trooster

Abstract The use of virtual worlds and serious gaming in education is growing. Data are limited on the efficacy and mechanism of action of these learning platforms. Therefore, we focus our research and development activities on this field. In our first study, we explored 12 educational initiatives in the virtual world Second Life (SL). There we found seven learning activities in SL with added didactic value: exercising skills, demonstrating learning topics, building, organizing events, organizing exhibitions, meeting people in order to learn and supervision of students. All these activities included aspects of social interaction and collaboration between students. Based on the experiences in these initiatives, we constructed a checklist for teachers considering the use of virtual worlds in education. With this checklist, we developed the serious game 3D-Trivial Pursuit in Second Life. In this game, students exercise exam questions for their education. In our second (pilot) study we found that the Trivial Pursuit game triggered the intrinsic motivation to learning in students, resulting in increased well-being and decreased drop-out of students. The results suggest that the synchronous communication between students during the game triggers social interaction, which is prerequisite for the observed intrinsic motivation. These examples show how our institute uses the results of our research for the development of new learning tools.

Keywords Serious gaming · Virtual worlds · Education

1 Introduction

The use of virtual worlds and serious gaming in education is growing. Data are limited on the efficacy and mechanism of action of these learning platforms. Therefore, we focus our research on this field. Using the virtual world Second Life

W. Trooster (✉)
Windesheim University Zwolle, Zwolle, The Netherlands
e-mail: W.Trooster@windesheim.nl

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_8,

we explore for which learning activities virtual worlds can be used with added value, and show how, with a checklist from our research, Serious Games (here Trivial Pursuit) can be designed.

Based on our research we consult teachers how to implement serious games and virtual worlds in their curricula, and train and help them to develop new learning tools of good quality. Finally, we study the use of these new learning tools in education in collaboration with our partners. In this paper we describe our activities.

2 Virtual World

In their (self) determination theory, Ryan and Deci (2000; see Fig. 1) indicate that in education intrinsic motivation of students is crucial for good learning results, well-being, and low drop-out of students. Prerequisite for this intrinsic motivation are three student-related factors: social interaction (student–student and student–teacher: "relatedness," "presence"), control of students on their learning process and competence to do the assignments.

Three platforms may trigger this intrinsic motivation: virtual worlds, serious games, and simulations. Simulations and serious games can be developed in virtual worlds. Therefore in the first study, we explored which learning activities in the virtual world Second Life (SL) have added didactic value for education. Using questionnaires we described 12 initiatives in higher education and secondary vocational education. Analysis of the data resulted in a list of seven learning activities in SL with added value: exercising skills, demonstrating learning topics, building, organizing events, organizing exhibitions, meeting people in order to learn, and supervision of students (see Fig. 2). All seven learning activities included aspects of social interaction and collaboration between students. In most SL-initiatives teachers reported better grades in comparison to traditional education. Those teachers reported increased intrinsic motivation in their students, and indicated as underlying student-factors social interaction and control on the learning process.

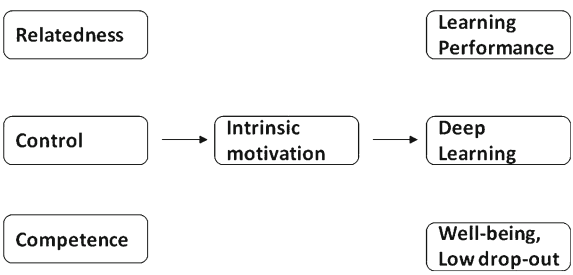
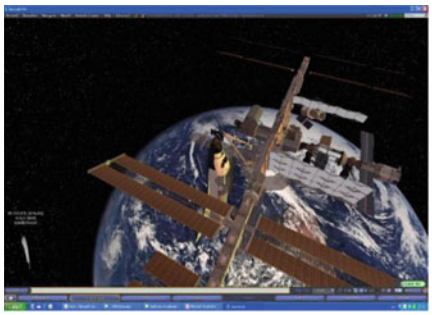


Fig. 1 The (self) determination theory by Ryan and Deci (2000)

Exercising skills



Demonstrating learning topics

Building

Organizing events

Organizing exhibitions

Meeting people in order to learn

Supervision of students

Fig. 2 Seven learning activities in SL with added value

Based on the experiences in these initiatives, we constructed a checklist for teachers considering the use of virtual worlds in education (see Table 1).

3 Serious Game

Using the checklist (Table 1) we developed the serious game 3D-Trivial Pursuit with exam questions (see Fig. 3). The objective of the game was to stimulate students to spend more time gaining knowledge that remains instantly available. The game has the same game rules as the original board game. Extra is the option to add sets of exam questions. With these educational questions, groups of students practice for their exams. During the game, students have the option to discuss about the exam questions using a synchronous communication function (chat) in the game. In our second (pilot) study, students in the control group individually used books gaining knowledge on the Dutch language. In the experimental group students for extra played the game with exam questions on the Dutch language, in groups of minimally two students. The results of this study show that students playing the game spent more time gaining knowledge. This is an indication that the game triggered the intrinsic motivation to learning in students. The results indicate that the game triggered social interaction between students (through synchronous communication), which is prerequisite for the observed intrinsic motivation. Beside increased intrinsic motivation, the students reported increased well-being and decreased drop-out. Likewise, the synchronous communication in the game could help to decrease the significant drop-out of students observed in distance

Table 1 Checklist for teachers considering the use of virtual worlds in education

Didactic factors
(1) Choose those learning activities for which the virtual world has added value compared to other media (e.g., the seven selected learning activities in this study)
(2) Make investments in adequate didactics:
a. Define and pursue the educational objectives of the course
b. Design adequate assignments
c. Organize adequate feedback and examination
d. Formulate transparent rules of conduct
e. Invest in adequate preparation for and teaching of the course
(3) Build/monitor/adjust the initiative in the virtual world
Organizational factors
(1) Create commitment (of teachers, students, and management (e.g., PR))
(2) Make connection to the policies of the institution
(3) Choose subgroups of students capable of working with the virtual world
(4) Minimize risks foreseen with the use of the virtual world
(5) Reduce fear that working with the virtual world is time-consuming
(6) Make investments in the development of expertise in the use of the virtual world
(7) Design protocols for the use of the learning activity in the virtual world
(8) Ensure support on IT

Fig. 3 The serious game 3D-trivial pursuit with exam questions

education, where traditionally communication is asynchronous. The number of students included in this study was to small ($n = 16$) to report on grades.

4 Recirculation of Knowledge From Research

With the knowledge from our research we consult teachers how to implement serious games and virtual worlds in their curricula, and train and help them to develop new learning tools of good quality. This way we helped to develop examples of all seven learning activities mentioned in the virtual world study, including more 3D-virtual boardgames and role-play games for education.

Finally, we plan to study the efficacy of these new learning tools in education.

5 Partners

Our institute collaborates with other institutes in the field of serious gaming to exchange techniques, didactics, and research facilities. In our role-play games we used primitive artificial intelligence to program nonhuman players. The interaction

of students with these "bots" was not natural. Therefore we started a collaboration with the institute TNO (Soesterberg, The Netherlands). With TNO we explored the use of their smart interface with high-tech artificial intelligence (Ashley) in our role-play-games. Furthermore, we collaborate with TNO in a study to explore the efficacy of five serious games in education.

With the Technical University of Twente (Enschede, The Netherlands) we collaborate in a Post-Master curriculum on the use of IT in Education. Serious Gaming is one of the topics in this curriculum.

Currently, we are in the midst of setting up a *Center on Serious Gaming in Education* at Windesheim University. In this center teachers and students are inspired (with seminars, consultancy, workshops) to use serious gaming in education. To facilitate that new learning materials are constructed, teachers with game concepts are matched with partners that can develop those games for them (internal partners (e.g., students), or external partners). Consultancy is provided to ensure that the products are didactically adequate learning materials. Research in this center is aimed at the efficacy of the developed learning materials and at the development of IT-competences of teachers

We are open to new partners interested in joining in on our Research & Development activities by exchanging techniques, didactics, and research facilities in order to optimize serious gaming in education.

6 Conclusion

In this chapter we described the merits of our research: we explored for which learning activities virtual worlds can be used with added value, and showed how with the checklist from our research Serious Games can be designed.

The limitation of the present study is that we used one virtual world (Second Life), and designed one Serious Game (Trivial Pursuit). In future work, we will expand our research to more virtual worlds and serious games.

References

1. Ryan RM, Deci EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 55(1):68–78

Serious Games for e-Health Care

Voravika Wattanasoontorn, Rubén Jesús García Hernández and Mateu Sbert

Abstract In this chapter, we present a state-of-the-art report on serious games dealing with e-health in a broad sense, including medicine, nursing, health care, and physical exercise. The games have been classified according to their main purpose (entertainment, teaching, or health), stages of the disease being treated and the type of users of the system (general population, patients, and health professionals). Additionally, 12 criteria dealing with the game technology have been selected for a fine-grain classification. Forty-one games from academic and commercial environments (including a variety of online games) have been described, analyzed, and classified.

Keywords Serious games · Healthcare

1 Introduction

In this chapter, we survey the serious games that are related to health and classify them with respect to different aspects. Not only games that have been described and evaluated in peer-reviewed publications are presented in this article but the scope of the survey also includes: (1) the commercial games (consoles and PCs),

V. Wattanasoontorn (✉) · R. J. G. Hernández · M. Sbert
Institute of Informatics and Applications, University of Girona, Girona, Spain
e-mail: voravika.wattanasoontorn@ima.udg.edu

R. J. G. Hernández
e-mail: rgarcia@ima.udg.edu

M. Sbert
e-mail: mateu@ima.udg.edu

V. Wattanasoontorn
Prince of Songkla University, Phuket, Thailand

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_9,

(2) online games, (3) games and application on mobile platforms, (4) games running on specialized platforms in clinics, hospitals, and patients homes.

We begin with reviews of important concepts. First, we provide an introduction of serious games, and then describe health. Next, the intersection of the two fields is described and provides a review on previous surveys. The details of the different ways in serious games in health will be classified. The next section contains a comprehensive list of serious games. A summary of their main characteristics, a comparison table, graphs are presented. Finally, discussion of the results and conclusion of the chapter are presented.

1.1 Serious Games

Today, the term serious game is becoming more and more popular even though there is currently no single definition of the concept. Serious games are defined in contraposition to entertainment games. They inherit gameplay characteristics from entertainment games, but the main focus may be learning or training and the lessons learnt are expected to be used in real-life work environments. Serious games are present in many areas of knowledge, including defense, manufacturing, education, and medicine, among others.

According to (Navarro et al. 2010), serious games are an emerging technology growing in importance for specialized training, taking advantage of 3D games and game engines in order to improve the realistic experience of users. We propose the characterization of serious games by using the following main components: rules and gameplay, challenges, interaction modes, and goals. The gameplay is the pattern defined through the game rules, which connects the player and the game. The goals may be explicit (stated as game objectives) or implicit. Implicit goals may include increasing skills and abilities, gaining knowledge, or acquiring experience (Fig. 1).

1.2 Importance of Health

The World Health Organization (WHO 2006) defined health in its broader sense as a "state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity." Other definitions simply require being free from illness or injury (Waite and Hawker 2009).

If we use the stricter definition, we may only consider games dealing with the different phases of illness development, both doctor training and patient familiarization with his illness. However, the use of the WHO definition allows us to consider a third variety of games, which has had a big success recently; games dealing with healthy habits such as exercise (including dancing and fitness games), so we shall use this broader definition in this chapter.

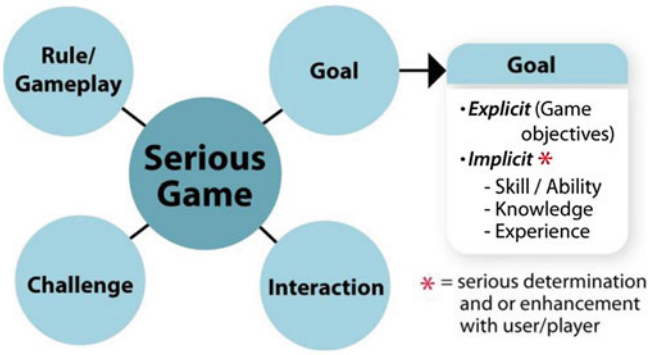


Fig. 1 The components of serious games

In Maslow's hierarchy of needs (Maslow and Frager 1987), health is represented in the second lowest level, after basic needs required for life are fulfilled. Health is then a very basic need, and maintaining health should therefore be a priority. Additionally, the desirable human characteristics located in higher levels of Maslow's pyramid, which are needed for a functioning, peaceful societies are negatively impacted by lack of health in the population. From the reasons stated above, Healthcare is one of the main issues that affects people the most in every stage of life (from infancy to old age). Many researches, such as (Kost 2001) and (Gostin 2000), have shown the need of highly trained and educated health care professionals to avoid medical errors, and the use of serious games in health can provide an additional mean to increase interest in training, education, and evaluation of their performance, as we will see in the next section.

1.3 The Use of Serious Games to Promote Health

Repetitive tasks are needed in many cases to treat patients, but patient boredom has a negative impact on the patient's willingness to continue the treatment. The use of tailored games to replace these tasks therefore has good results.

Additionally, since the recent explosion of videogames, which now are used in two-thirds of households by people of all ages (Online ED 2012), patients can feel more at ease and enjoy their treatment performing an activity they like.

Since 2002 (Roubidoux 2009), many serious games in the field of e-health have been developed, dealing with a wide variety of aspects of surgeon training, radiology operation, Cardiopulmonary Resuscitation (CPR), and patient care, among others. Games aimed at patients have also been developed.

The previous surveys of serious games are described next; (Lopes and Bidarra 2011) present the state of adaptivity in general games and simulations focusing on the purposes, targets, and methods from both academia and industry. Bartolome

et al. (2011) present a systematic review of 21 serious games for health and education described at scientific papers written in English from 2003 to 2011 and projects from the 7th Framework Program. Kato (2010) summarizes the scientific literature of commercially available and tailor-made games used for education and training with patients and medical students and doctors. The classification is based on diseases. Rego et al. (2010) propose a classification designed to properly distinguish and compare eight serious games for rehabilitation systems with respect to their fundamental characteristics. They also describe a particular serious game for rehabilitation, RehaCom, as a case study. Watters et al. (2006) explore the use of games for children with long-term treatment regimes, where motivation for compliance is a key factor in the success of the treatment.

2 Classification of Surveyed Serious Games for Health

There are many interesting criteria to classify serious games in health. In this section, we describe these criteria in detail (Fig. 2).

2.1 Classification by Main Purpose

There are three main purposes in serious games for health:

- The main purpose is entertainment but there is a need to move some parts of the body so the wellness is gotten as a bonus, such as dance–dance revolution (DDRGame 2012), which is the pioneering series of the rhythm and dance genre in video games. The exercise (commercial) games became famous, since Wii was released by Nintendo on November 19, 2012. The motion controlled over the avatar by various accessories inspired people to exercise with the video games.
- The main goal is health but the game is used as a tool to pass on knowledge and/or skills. To use the capability of the game engine, various serious health contents are conveyed to players; for example, Fatworld (ITVS's Electric Shadows 2007), Re-Mission (Tate et al. 2009), Air Medic Sky 1 (University

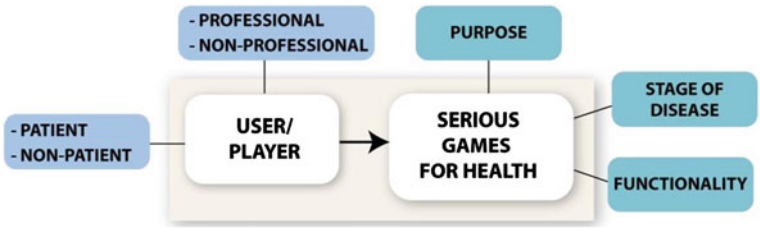


Fig. 2 Classification of serious games for health

Medical Center Utrecht 2012), and many other games shown in Sect. 3 (Review Studies).
- Serious use in health and medical purpose but with a need to simulate the situation to avoid/concern risk, safety, budget, etc. Most of the games in this category are simulation games with use in health and medical field such as Virtual Dental Implant Training Simulation Program (BreakAway, Ltd. 2010), EMSAVE (Emergency Medical Services for the disAbled Virtual Environment) (Vidani et al. 2010), Olive: 3D Hospital Training (ScienceApplications 2010), etc.

2.2 Classification by Player (Patient/Non-Patient)

2.2.1 Patient

To classify the health objectives related to stage of disease (Table 1)

- Health monitoring
- Detection
- Treatment
- Rehabilitation
- Education for self/directed care

2.2.2 Non-Patient

- Health and wellness: Focuses more on lifestyle issues and their relationships with functional health; data from the Alameda County Study (Table 2) (Housman and Dorman 2005) suggested that people can improve their health via (1) exercise, (2) enough sleep, (3) maintaining a healthy body weight, (4) limiting alcohol use, and (5) avoiding smoking.
- Training and Simulation: For both professional and non-professional.

Table 1 Classification of serious games for health (for patient)

Game purpose	Serious games
Monitoring	CHF telemanagement systems (Finkelstein et al. 2010), Healthcare monitoring (Fergus et al. 2009) and The U-health monitoring system (Lee et al. 2009)
Detection	Unobtrusive health (Mckanna et al. 2009) and EEG-based serious games (Wang et al. 2010)
Treatment	Match-3 (Scarle et al. 2011), diagnosis and management of parkinson (Atkinson and Narasimhan 2010), and social skills (Bartolome et al. 2010)
Rehabilitation	Neuropsychological rehabilitation (Grau et al. 2010), chronic pain rehabilitation (Schnauer 2011), upper limb rehabilitation following stroke (Burke 2009), and after parkinson's disease (Red Hill Studios, Inc. 2011)
Education	Re-Mission (Tate et al. 2009)

Table 2 Classification of serious games for health (for non-patient)

Game purpose	Serious games
Health and wellness	A sensory gate-ball game (Kim et al. 2009), dancing in the streets (DITS) (Clawson et al. 2010), fitness adventure (Laikari 2009), Virku (Vtnen and Leikas 2009), and MoFun circus (Nordic Innovation Center 2007c)
Training: professional	HumanSim (Preview) (Applied Research Associates, Inc. 2012), virtual dental implant training Simulation Program (BreakAway, Ltd. 2010), nursing and midwifery (Skills2Learn, Ltd. 2010), pulse (Eliane 2007), EMSAVE (Emergency Medical Services for the disabled) (Vidani et al. 2010), MUVE market virtual patient CareSimulation Lab (John 2010), Olive: 3D hospital training (Science Applications 2010), game-based learning for virtual patients (SAIC, Inc. and The Faculty of Medicine at Imperial College London 2008), Nurse education (AndyWBlackburn 2008), Virtual patient (ScienceDaily 2009), medical simulation training program (JDoc) (Sliney and Murphy 2008), VI-MED (Mili et al. 2008), (Sabri et al. 2010), and Air Medic Sky 1 (University Medical Center Utrecht 2012)
Training: non-professional	Terveellinen Ateria (Janomedia 2006), MC Urho (Nordic Innovation Center 2007b), Valion Energiasummaaja (Nordic Innovation Center 2007a), Fatworld (ITVS's Electric Shadows 2007), and the food detectives fight BAC! game (The Partnership for Food Safety Education 2008)

2.3 Classification by Stage of Disease

Following Merrill (2010), we can classify the progression of an illness in the following stages (Table 3):

2.3.1 Stage of Susceptibility

This is the first stage, in which the person is still healthy. Nevertheless, some people have a genetic predisposition to develop certain illnesses later in life, so even in this healthy stage it is advisable to perform periodic checkups. A serious game in this stage helps the user to familiarize with monitoring procedures, and with illnesses that he or she will possibly develop later in life. Environmental risks or unhealthy surroundings can also affect the probability of developing illnesses, so this should also be taken into account.

Table 3 The objective of the serious games according to stage of disease

Stage of disease	Serious game purpose
Stage of susceptibility	Monitor
Pre-symtomatic stage	Detect
Stage of clinical disease	Treatment and therapy
Stage of disability rehabilitation	Track and trace

2.3.2 Pre-Symptomatic Stage

In this stage, people still feel healthy although the illness is already present. For example, the number of viric particles may still be too small to produce a response in the body, or a failing organ may still be able to cope with the added pressure of the illness with no external indications. The beginning of this stage may be discovered by the periodic checkups mentioned above. In this stage, the specific illness is now known, and the chances of developing it are very high, so more focused serious games can be used to show the patient, the relevant aspects of his illness and his treatment.

2.3.3 Stage of Clinical Disease

In this third stage, the symptoms of the illness are already manifest in the patient. If the illness was not detected in the previous stages, serious games can be used to familiarize the patients with the expected progression of their illness, and the treatment procedures. Alternative treatments can also be shown using games. Games intended to be played by doctors or other medical staff, usually focus on this stage as well.

2.3.4 Stage of Recovery, Disability, or Death

In this last stage of the illness, three different outcomes are possible: the illness may be cured, returning the patient to health or to another stage of susceptibility, or it may have serious effects on the patient's health, making them unable to function at previous levels (disability). The worst outcome of an illness, of course, is the death of the patient. Serious games in this stage normally deal with the rehabilitation procedure, or help the patient cope with their disabilities.

2.4 Classification by Functionality

Rego et al. (2010) identify some criteria for the classification of serious games for health (Fig. 3). We build upon their criteria and add some more interesting characteristics. The description of our classification system follows:

- Application area: The application area or domain describes the part of the real world being modeled by the software. In serious games for health, we will distinguish two main aspects: cognitive skills such as memory, attention span, concentration and reasoning (Cog), and motor skills such as general coordination or re-learning to walk after injuries (Mot).

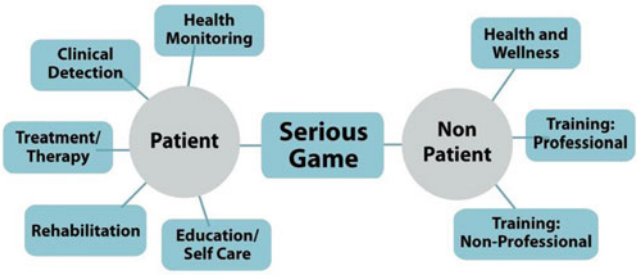


Fig. 3 Classification of serious games for health by player

- Interaction Technology: Interaction technologies are the different paradigms for establishing communication between humans and computers. Both hardware and software interfaces are included. Traditionally, mouse and keyboards have been used; newer means of interaction include Virtual Reality (using head-mounted displays (HMDs), computer monitors, haptic or pseudo-haptic devices such as gloves or pens, or tracking devices. Webcams and web applications are also common. Patients can affect virtual objects in real-time using a variety of senses (vision, hearing, and touching).
- Game Interface: The virtual world inside the game can simulate the real world (three dimensional games, 3D) or provide a top–bottom or sideways perspective of a simpler world (two dimensional games, 2D).
- Number of Players: The number of users concurrently using the world of the game. In general, we distinguish single player games (for one person) and multiplayer games (for two or more people).
- Game Genre: Games can be categorized according to their gameplay; we can distinguish adventure, strategy, simulation, sports and puzzles, among others. In games for health, the games, which evaluate coordination and the movement, are common in rehabilitation; other genres are also used for different tasks.
- Adaptability (Yes/No): Old games used to have a fixed level of difficulty, which could either be programmed or chosen before the game started. Nowadays, many games try to adapt their difficulty to the skill of the player, in order to increase playability and enjoyment by the user. In e-health, adaptability is an excellent trait because it allows the patient to test his limits in a controlled manner.
- Performance Feedback (Yes/No): Performance feedbacks are the indications of the game dealing with showing the users their status and abilities. They allow patients to feel confident that they are progressing and to detect and fix their failures. The feedback can be audition, visual, or haptic.
- Progress monitoring: For patient evaluation, having logs of the patient actions inside the game can be an invaluable asset. We call this feature progress monitoring, since it allows the doctors to monitor the progress of the patients as a function of time.

- Game portability: Portability refers to being able to physically move the game hardware. In particular, we distinguish games located at a hospital or clinic, and games which can be used at home or which can be held portably by the user.
- Game Engine: A game engine is a platform, which provides commonly used functionality in games, so that the game developers can focus on higher level of game design and functionality. The engine provides an API to access lower level functionality and a set of predefined models and materials and scenes (Cheng-yong and Wei-ming 2010).
- Platform: The hardware the game is run on. This may include personal computers (PCs), commercial game consoles (Nintendo Wii, Microsoft Xbox), portable consoles, or custom hardware.
- Health Objective: Objectives related to health include monitoring, detection and treatment of illnesses, rehabilitation, education, health, and wellness, training for patient, and training for non-patient so that they can take care of themselves, or at least know and understand the caring procedures carried upon them. Other general objectives include healthy habits such as exercise, sleeping patterns, and making patients avoid excessive alcohol intake and smoking.
- Connectivity: Games might require an internet or network connection (online) or they may be played in standalone computers (offline).

3 Review Studies

Several serious games and applications for health have been reported in literature. In this section, we review the work developed in this area, including additional games from commercial or online sources. We provide a short summary of each game, indicating their most relevant characteristic, and finish the section with a table overviewing the characteristics of all the games according to the different aspects detailed in the previous section. The games studied are:

- Game1—CHF telemanagement systems (Finkelstein et al. 2010): Home Automated Tele-management (HAT) system for chronic disease management in the patient home. The system helps patients with congestive heart failure (CHF) monitor their symptoms, weight changes, and quality of life while teaching the patient, the characteristics of their disease. The system runs on the Nintendo Wii console. An internet connection is required.
- Game2—Healthcare monitoring in the home (Fergus et al. 2009): A medical diagnostic gaming environment that is used to gather patient information in a casual, non-obtrusive manner that is relaxing for the patient.
- Game3—The u-health monitoring system, with a Nintendo DS (Lee et al. 2009): This application displays the bio-signals of patient onto a monitor of personal computer and LCD of a Nintendo DS using a bio-signal measurement device connected via wireless protocol.

- Game4—Unobtrusive Health (Mckanna et al. 2009): 21 Tally is a collection of 2D games used to detect divided attention unobtrusively, by using performance on a computer game designed to force players to attend to different dimensions simultaneously in order to succeed.
- Game5—EEG-Based Serious Games (Wang et al. 2010): The EEG-based concentration games named Brain Chi (2D) and Dancing Robot (3D) were developed for concentration level control.
- Game6—Match-3 (Scarle et al. 2011): A serious game designed to combat childhood obesity. The Wii-mote is being used for a rowing action which propelled the coracle forward, while direction is altered by leaning left and right on the Wii-fit.
- Game7—Medical Gaming Environment for Diagnosis and Management of Parkinson's Disease (Atkinson and Narasimhan 2010): A medical diagnostic gaming environment that is used to gather patient information in a casual, non-obtrusive manner that is relaxing for the patient. The system employs the Novint Falcon human interface device (Novint Technology, Inc. 2011) to guide a patient who may have problems reaching a specified goal within the game.
- Game8—Neurological development (Bartolome et al. 2010) present a 3D game that can analyze the behavior and promote certain social skills (conversation, negotiation, etc.) of people with neurological development disabilities. The treatment has been planned at three levels. Each level treats a characteristic related to socialization, integration, and the expression of feelings.
- Game9—Neuropsychological Rehabilitation (Grau et al. 2010): single-user and first-person-shooter tasks. Patients navigate through the virtual environment and perform cognitive tasks.
- Game10—Chronic Pain Rehabilitation (Schnauer 2011): This system provides multimodal interaction including full body motion capture by the use of Microsoft Kinect, and other bio-signal capture devices. The patients can manage their state and train physically on their own.
- Game11—Serious Games for Upper Limb Rehabilitation Following Stroke (Burke 2009): Several systems are developed by research group at the University of Ulster, for upper limb stroke rehabilitation through the integration of 3D virtual environments and sensor and camera technology. There are five games including:
  - Game11a: Catch task for bilateral rehabilitation.
  - Game11b: Adaptive whack a mouse game, which designed to encourage movement and to improve the accuracy and speed of the users upper limb movement.
  - Game11c: Rabbit Chase developed for single arm rehabilitation (either right or left arm).
  - Game11d: Arrow Attack for bimanual rehabilitation (both arms) and
  - Game11e: Virtual vibraphone, the use of Nintendo Wii remote controllers for wrist and arm rehabilitation.

- Game12—Games for people with Parkinson's Disease (Atkinson 2010, Red Hill Studios 2011): These nine therapeutic games by Red Hill Studios and the School of Nursing at the University of California San Francisco can help Parkinson's disease patients increase their balance. The games are played by performing movements, which are known to be beneficial for balance control and the movements, then are captured and processed by the system. Patients can use this immersive and engaging virtual world to practice gait and balance in a more interesting setting.
- Game13—Re-Mission (Tate et al. 2009): A video game with 20 levels that takes the player on a journey through the body of young patients with different kinds of cancer, released by the non-profit HopeLab. The main aim is to engage young cancer patients through entertaining game play while impacting specific psychological and behavioral outcomes associated with successful cancer treatment.
- Game14—A Sensory Gate-Ball Game (Kim et al. 2009): PC-based 3D graphics game design for aged people, uses a realistic gate-ball stick and balls as interfaces. In the game, players use the same stick and ball as the real gate-ball.
- Game15—Dancing in the Streets (DITS) (Clawson et al. 2010): DITS is a mobile phone version of the popular arcade game Dance, Dance, Revolution-TM(DDR). Instead of using a dance pad, DITS uses wireless 3-axis accelerometers that are worn around the player ankles and uses a mobile phone to control the game and to display graphics.
- Game16—Fitness Adventure (Laikari 2009): This application platform, which is able to take advantage of a variety of mobile phones, location information and bluetooth GPS receivers, combines mobile games with exercising outdoors. The end result is a location-aware fitness game.
- Game17—Virku-Virtual Fitness Center (Vtnen and Leikas 2009): This system allows users to exercise in a virtual environment. The game is controlled by a user interface based on an exercise cycle, and users may practice individually or in a group. The virtual world affects the difficulty of the exercises in a coherent manner, for example, changing the resistance of bicycles pedals during virtual hill climbing.
- Game18—MoFun Circus (Nordic Innovation Center 2007c): This installation, located at Heureka (Vantaa, Finland) is a cooperative multiplayer action game. After users meet in a circle, a trampoline is created, and users move and capture falling objects. A camera is used to follow the users and display the activity on-screen.
- Game19—HumanSim (Preview) (Applied Research Associates, Inc. 2012): This preview shows the appearance of the HumanSim immersive world. In this world, doctors and nurses train to learn the nuances of complex, unusual, or other error-prone tasks until they become experts. A high-quality virtual hospital is modeled, including operating rooms and other important spaces, and populated with doctors and nurses. It runs on an Apple iPad.
- Game20—Virtual Dental Implant Training Simulation Program (BreakAway, Ltd. 2010) (Medical College of Georgia School of Dentistry faculty 2009): This software provides a 3D virtual environment for students to train in the correct

decision-making protocol to determine patient preparation (both physical and mental) for dental implant surgery. After ensuring that surgery is needed, the student can practice the procedure in a safe and realistic environment.

- Game21—Nursing and Midwifery (Skills2Learn, Ltd. 2010): This program helps nurses and midwives increase their ability to assess patients. The interactive scenario is based on the simulation of the 36 weeks of pregnancy realistically. Multimedia and virtual reality are combined, so that the user can move inside the hospital, interact with patients, and perform the needed tests with the correct instruments.
- Game22—Pulse!! The Virtual Clinical Learning Lab (Eliane 2007): This is the first immersive virtual learning space where health care professionals can train their clinical skills. State-of-the-art graphics create a virtual world in which both civilians and military professionals in the field of health care can practice clinical skills in an interactive and believable universe. This allows them to respond better to patients with injuries due to combat, bioterrorism, or other catastrophes.
- Game23—EMSAVE (Vidani 2010): is a system for training in emergency medical procedures concerning disabled patients. It allows users to experience emergency situations involving disabled persons.
- Game24—MUVE/Market Virtual Patient Care Simulation Lab (John 2010): This software can be used to create simulations of patents useful in training students and professionals (nurses, pharmacists, paramedics, emergency medical technicians, social workers, etc.). The environment is part of Second Life: a massive multiplayer online role-play game (MMORPG) that allows users to choose a different lifestyle.
- Game25—Olive: 3D Hospital Training (ScienceApplications 2010): The On-Line Interactive Virtual Environment (OLIVE) is a dynamic software platform which eases the development and deployment of collaborative virtual worlds. The worlds can be customized and privacy and security are taken into account in the platform. In the virtual world, users can perform planning, training, rehearsing, and operating over long distances using computer networks. The technology has been used to train different scenarios in a hospital operating theater. Real nurses, doctors, and patients appear in the game using their avatars, and remote users can receive standardized Sharable Content Object Reference Model (SCORM) training data and participate in the virtual world.
- Game26—Game-based learning for Virtual Patients–Multi patients (SAIC Inc and The Faculty of Medicine at Imperial College London 2008). A region in Second Life created by the Faculty of Medicine at Imperial College London provides a learning space where virtual patients suffering five different respiratory illnesses (such as lung cancer or pneumonia) can be diagnosed, investigated, and treated by players wanting to perform role-playing learning activities under the feedback and guidance of medical staff.
- Game27—Nurse education (AndyWBlackburn 2008): A virtual learning environment (Second Life), for use in nurse education. Developed at Glasgow Caledonian University.

- Game28—Virtual Patient (ScienceDaily 2009) [Game28]: This system, developed by Keele University, trains pharmacists by using a virtual patient. Traits such as race, age, and gender (which sometimes affect responses to procedures and medications) are taken into account in the treatment of patients, so that learners can understand their clinical significance. Dyspepsia and hypertension are examples of the possible illnesses included in the system.
- Game29—Medical Simulation Training Program (JDoc) (Sliney and Murphy 2008): This system provides a computer-aided junior doctor simulator, using both 1st and 3rd person perspectives to produce immersion. It can be used for training and teaching junior doctors their interpersonal, communication and decision-making skills, and to ease the transference of the medical information available to them.
- Game30—VI-MED (Mili et al. 2008): A virtual training to be used as a precursor and as a supplement to real practical training.
- Game31—(Sabri et al. 2010): Presented an interactive, multiplayer serious game for the purpose of training cardiac surgeons/residents the series of steps comprising the Off-Pump Coronary Artery Bypass (OPCAB) surgical procedure.
- Game32—Air Medic Sky 1 (University Medical Center Utrecht 2012): This interactive biofeedback game consists of mini-games and lectures to describe the basic concepts required for efficient communication and teamwork resulting in patient safety. Complex situations are presented which mirror common occurrences in the work of junior doctors. The solving of these solutions allows doctors to gain insights in how physiological functions affect performance.
- Game33—Terveellinen Ateria (Janomedia 2006): This interactive program aids practical nurses and comprehensive school staff train in the practical aspects of preparing meals for people with different nutritional requirements. The Finnish plate model is used as a basis, and it can be applied to solve the dietary needs of people. The energy content of meals is calculated according to the foods included and their size. Teaching healthy nutrition and weight management habits to the general population is one objective of the program.
- Game34—MC Urho (Nordic Innovation Center 2007b): This system contains a plethora of information regarding lifestyle effects on health. The game can be used in biology and health education classes to teach young people about the effects of smoking, high blood pressure and cholesterol.
- Game35—Valion Energiasummaaja (Nordic Innovation Center 2007a): This online game, aimed at both children and adults, helps build a healthy and balanced breakfast (or general snacks). Food is chosen from a variety of options in a virtual refrigerator, and dragged into a plate. The effects of the meal on blood sugar are shown, and possible improvements in the meal are suggested.
- Game36—Fatworld (ITVS's Electric Shadows 2007): This videogame explores the relationships between nutrition, obesity and socioeconomic factors in the contemporary U.S. Budgets, subsidies, regulations, and physical world characteristics are taken into account.

Table 4 Classification and comparison of health games (1)

	Application area	Interactive technology	Interface	Players	Genre	Adaptability	Progress Monitoring	Performance. Feedback	Portability	Engine	Platform	Objective	Connectivity
[Game1]	Mot	Wiimote	2D	Single	App	x	•	•	•	Flash	Wii	MON	On
[Game2]	Mot	WSN	2D	Single	App	x	•	•	x	–	PC	MON	On
[Game3]	Mot	Electrode	2D	Single	App	x	•	•	•	PALIB	DS	MON	On
[Game4]	Cog	Mouse	2D	Single	Puzzle	•	•	x	•	–	PC	DET	Off
[Game5]	Cog	EEG	2D	Single	Action	–	–	–	x	–	PC	DET	Off
[Game6]	Mot	Wii Fit	3D	Single	Adventure	•	–	•	•	–	Wii	TRT	Off
[Game7]	Cog	Novint	2D	Single	Action	•	–	–	•	SDK	PC	TRT	Off
[Game8]	Cog	Wiimote	3D	Single	Role play	•	•	•	•	Director	PC	TRT	Off
[Game9]	Cog	Mouse	3D	Single	Role play	•	–	•	•	–	PC	REH	Off
[Game10]	Mot	Kinect	3D	Single	Adventure	•	•	•	x	Unity3D	PC	REH	Off
[Game11a]	Mot	VR	3D	Single	Action	•	•	•	•	OGRE	PC	REH	Off
[Game11b]	Mot	HMD	3D	Single	Action	•	•	•	•	–	PC	REH	Off
[Game11c]	Mot	Webcam	2D	Single	Action	•	•	•	•	XNA	PC	REH	Off
[Game11d]	Mot	Webcam	2D	Single	Action	•	•	•	•	XNA	PC	REH	Off
[Game11e]	Mot	Wiimote	2D	Multi	Action	•	•	•	•	–	PC	REH	Off
[Game12]	Mot	–	2D/3D	–	Mix	–	–	–	–	–	PC	REH	–
[Game13]	Cog	Mouse	3D	Single	Adventure	•	•	•	•	–	PC	EDU	Off
[Game14]	Mot	Sensor	3D	Single	Sport	•	•	•	x	–	PC	H&W	Off
[Game15]	Mot	Sensor	2D	Single	Exergame	x	•	x	•	–	Mobile	H&W	Off
[Game16]	Mot	RFID	2D	Single	Exergame	•	•	•	•	SMAC	Mobile	H&W	On
[Game17]	Mot	Cycle	3D	Single	Exergame	•	•	–	x	–	PC	H&W	Off
[Game18]	Mot	Camera	2D	Multi	Action	–	•	–	x	–	PC	H&W	Off
[Game19]	Cog	Touch	3D	Single	Simulation	•	•	•	•	Unreal	iPad	P	Off
[Game20]	Cog	Mouse	3D	Single	Simulation	•	•	•	•	BreakAway	PC	P	Off
[Game21]	Cog	Mouse	3D	Single	Simulation	•	•	•	•	–	PC	P	Off
[Game22]	Cog	Mouse	3D	Single	Simulation	•	•	•	•	BreakAway	PC	P	Off
[Game23]	Cog	Mouse	3D	Single	Simulation	•	•	•	•	NeoAxis	PC	P	Off
[Game24]	Cog	Mouse	3D	Single	Simulation	•	•	•	•	–	PC	P	On
[Game25]	Cog	Mouse	3D	Multi	Simulation	•	•	•	•	–	PC	P	On

Table 9.5 Classification and Comparison of Health Games (2)

	Application area	Interactive technology	Interface	Players	Genre	Adaptability	Progress Monitoring	Performance. Feedback	Portability	Engine	Platform	Objective	Connectivity
[Game26]	Cog	Mouse	3D	Single	Simulation	•	•	•	•	SecondLife	PC	P	On
[Game27]	Cog	Mouse	3D	Single	Simulation	–	–	–	•	SecondLife	PC	P	On
[Game28]	Cog	VoicRe	3D	Single	Simulation	•	•	•	•	–	PC	P	On
[Game29]	Cog	Mouse	3D	Single	Simulation	•	•	•	•	Torque	PC	P	On
[Game30]	Cog	Mouse	3D	Single	Simulation	•	•	•	•	–	PC	P	On
[Game31]	Cog	Mouse	3D	Multi	Simulation	•	•	•	•	In House	PC	P	On
[Game32]	Cog	Biof B.	3D	Multi	Simulation	•	•	•	•	–	PC	P	On
[Game33]	Cog	Mouse	2D	Single	Puzzle	•	•	–	•	Flash	PC	NP	On
[Game34]	Cog	Mouse	2D	Single	Quiz	•	x	–	•	Flash	PC	NP	On
[Game35]	Cog	Mouse	2D	Single	Puzzle	•	•	–	•	Flash	PC	NP	On
[Game36]	Cog	Mouse	2D	Single	Role play	•	•	•	•	Flash	PC	NP	Off
[Game37]	Cog	Mouse	2D	Single	Puzzle	•	•	x	•	Flash	PC	NP	On

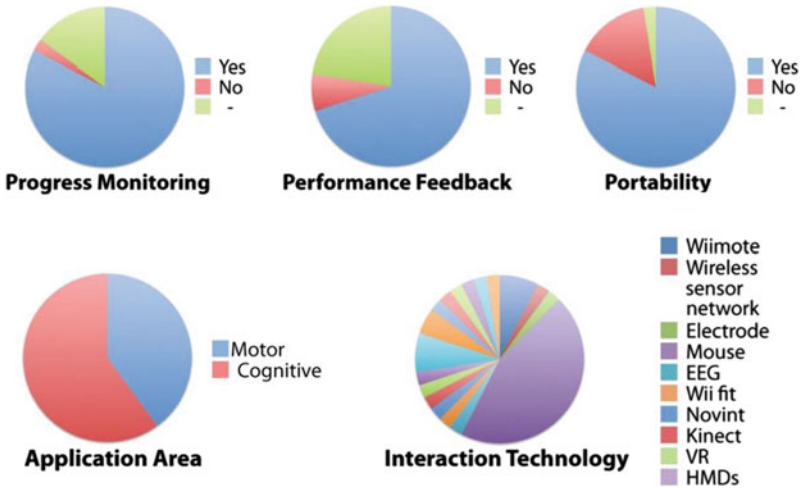


Fig. 4 Breakdown of the values of each characteristic of serious games for health present in our survey (1)

- Game37—The Food Detectives Fight BAC! (The Partnership for Food Safety Education 2008): A 2D web base game for 8–12 year old children to learn about foodborne illness. Created by The Partnership for Food Safety Education.

4 Discussion

Tables 4 and 5 display the classification and comparison of health games. The dash means that this feature is not mentioned in the bibliographic reference, which described the game. Figures 4 and 5 provide a graphical summary of the data.

The objectives are quite varied, but an emphasis can be seen on professional users and rehabilitation. We can see that both motor and cognitive abilities are well-represented application areas. Progress monitoring, performance feed and portability are important features, present in the majority of the games. We can see that adaptability is a useful characteristic, present in about 3/4 of the games.

There are a few used platforms, but most of the games have been designed to run on PCs, and internet connection is used in almost half of the games surveyed. Interaction technologies are quite varied, although the standard mouse interface is used in almost half of the games. There are a wide variety of engines used, although this is not a reported characteristic. Flash is the engine used most often, by a wide margin, and most of the games are currently single player. The simulation and action genres dominate, although there are large varieties of other genres, and 2D and 3D interfaces are both well represented in the sampled games.

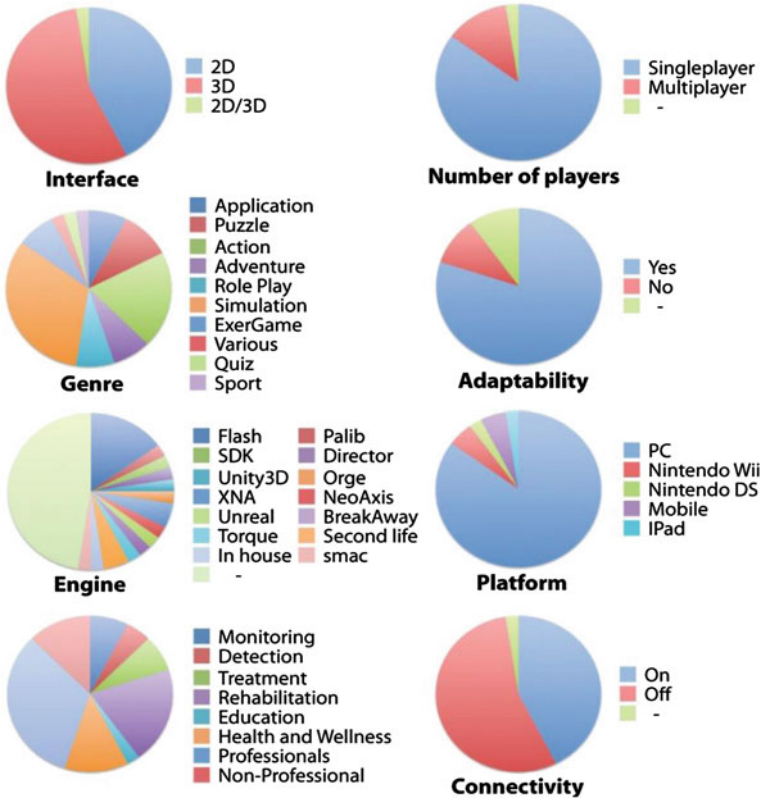


Fig. 5 Breakdown of the values of each characteristic of serious games for health present in our survey (2)

5 Conclusion

We have studied 40 serious games from different areas of health and well being, classifying them according to their characteristics, and described the most commonly present ones. The typical game can be summarized as a portable PC game programmed in Flash, using mouse interaction and including progress monitoring, performance feed and adaptability, although the variability of games is quite large in many aspects. For the future improvement, authors plan to explore more serious games and applications for mobile devices such as iPhone, iPad, etc.

Acknowledgments This work has been supported by the research project coded TIN2010-21089-C03-01 (Spanish Commission for Science and Technology), by grant 2009SGR643 (Catalan Government) and by a mobility grant from the Prince of Songkla University, Thailand.

References

Alhadeff E (2007) Pulse!! news: serious games field testing begins. http://futuremakingseriousgames.blogspot.com/2007/02/pulse-news-serious-games-field-testing.html Accessed 7 March 2012

AndyWBlackburn (2008) Nurse education in second life. Glasgow Caledonian University. http://www.youtube.com/watch?v=xidko60S2Uk&feature=related Accessed 7 March 2012

Applied Research Associates, Inc (2012) HumanSim: a high-fidelity virtual hospital. http://www.humansim.com Accessed 6 March 2012

Atkinson SD, Narasimhan VL (2010) Design of an introductory medical gaming environment for diagnosis and management of Parkinson's disease. Trendz in Information Sciences & Computing (TISC), pp 94–102, 17–19 Dec. 2010

Bartolome NA, Zorrilla AM, Zapirain BG (2011) Can game-based therapies be trusted? Is game-based education effective? A systematic review of the serious games for health and education. 16th international conference on Computer Games (CGAMES), pp 275–282, 27–30 July 2011

Bartolome NA, Zorrilla AM, Zapirain BG (2010) A serious game to improve human relationships in patients with neuro-psychological disorders. International IEEE consumer electronics society's games innovations conference (ICE-GIC), pp 1–5, 21–23 Dec. 2010

BreakAway, Ltd (2010) Serious games for healthcare markets. http://www.breakawaygames.com/serious-games/solutions/healthcare/ Accessed 6 March 2012

Burke JW, McNeill MDJ, Charles DK, Morrow PJ, Crosbie JH, McDonough SM (2009) Serious games for upper limb rehabilitation following stroke, VS-GAMES '09. Conference in games and virtual worlds for serious applications, pp 103–110, 23–24 March 2009

Cheng-yong X, Wei-ming X (2010) Constructing 3d game engine based on XNA. Comput Knowl Technol 6:3401–3402

Clawson J, Patel N, Starner T (2010) Dancing in the streets: the design and evaluation of a wearable health game. International symposium on wearable computers (ISWC), pp 1–4, 10–13 Oct. 2010

Education Database Online (2012) Videogame statistics. http://www.onlineeducation.net/ Accessed 30 March 2012

Fergus P, Kifayat K, Cooper S, Merabti M, El Rhalibi A (2009) A framework for physical health improvement using wireless sensor networks and gaming, Pervasive Health 2009. 3rd international conference on pervasive computing technologies for healthcare, pp 1–4, 1–3 April 2009

Finkelstein J, Wood J, Cha E, Orlov A, Dennison C (2010) Feasibility of congestive heart failure telemanagement using a wii-based telecare platform. Annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 2211–2214, 31 Aug.–4 Sept. 2010

Gostin L (2000) A public health approach to reducing error: medical malpractice as a barrier. JAMA 283(13):1742–3, url http://www.biomedsearch.com/nih/public-health-approach-to-reducing/10755503.html

Grau S, Tost D, Campeny R, Moya S, Ruiz M (2010) Design of 3d virtual neuropsychological rehabilitation activities. 2nd international conference on games and virtual worlds for serious applications pp 109–116

HCI Lab, University of Udine (2011) The EMSAVE System. http://hcilab.uniud.it/soccorsodisabili/results.html Accessed 7 March 2012

Housman J, Dorman S (2005) The alameda county study: a systematic, chronological review. American J Health Edu 36:302–308

ITVS's Electric Shadows (2007) Fatworld. http://fatworld.org/ Accessed 26 March 2012

Janomedia (2006) Terveellinen Ateria. http://www03.edu.fi/oppimateriaalit/healthy_meal/ Accessed 26 March 2012

Kato PM (2010) Video games in health care: closing the gap. Rev Gen Psychol 14:113–121

Kim J-A, Kang K-K, Yang H-R, Kim D (2009) A sensory gate-ball game for the aged people and its user interface design. Conference in games and virtual worlds for serious applications, pp 111–116
Kost G (2001) Preventing medical errors in point-of-care testing: security, validation, safeguards, and connectivity. Arch Pathol Lab Med 125(10):1307–1315
Laikari A (2009) Exergaming-gaming for health: a bridge between real world and virtual communities. IEEE 13th international symposium on consumer electronics (ISCE), pp 665–668
Lee S, Kim J, Kim J, Lee M (2009) A design of the u-health monitoring system using a nintendo ds game machine. IEEE Eng Med Biol Soc pp 1695–1698
Lopes R, Bidarra R (2011) Adaptivity challenges in games and simulations: survey. IEEE Trans Comput Intell AI Games 3:85–99
Maslow A, Frager R (1987) Motivation and personality. Harper and Row, New York http://books.google.es/books?id=L7_uAAAAMAAJ
Mckanna JA, Jimison H, Pavel M (2009) Divided attention in computer game play: analysis utilizing unobtrusive health monitoring. 31st annual international conference of the IEEE EMBS, pp 6247–6250, Sept. 2009
Medical College of Georgia School of Dentistry faculty (2009) Simulation helps students learn dental implant procedures. http://news.georgiahealth.edu/archives/1921. Accessed 6 March 2012
Merrill R (2010) Introduction to epidemiology. Jones and Bartlett Publishers, London http://books.google.es/books?id=RMDBh6gw1_UC
Mili F, Barr J, Harris M, Pittiglio L (2008) Nursing training: 3d game with learning objectives. 1st international conference on advances in computer-human interaction, pp 236–242
Miller J (2010) MUVE market virtual patient care simulation Lab. http://www.youtube.com/watch?v=FWUpXar6sh8. Accessed 7 March 2012
Navarro A, Pradilla JV, Madrinan P, (2010) Work in progress serious 3d game for mobile networks planning. Frontiers in education conference (FIE), 2010 IEEE, pp T1F-1,T1F-2, 27–30 Oct. 2010
Nintendo, Inc (2012) Wii. http://www.nintendo.com/wii/. Accessed 29 March 2012
Nordic Innovation Centre (2007a) Energiasummaaja. http://nsg.jyu.fi/index.php/Energiasummaaja. Accessed 26 March 2012
Nordic Innovation Centre (2007b) MC Urho. http://nsg.jyu.fi/index.php/MC_Urho/. Accessed 26 March 2012
Nordic Innovation Centre (2007c) MoFun Circus. http://nsg.jyu.fi/index.php/MoFun_Circus. Accessed 26 March 2012
Novint Technologies, Inc (2011) The most immersive way to play video games. http://www.novint.com/index.php/novintfalcon. Accessed 29 March 2012
Red Hill Studios (2011) Games for people with Parkinson's Disease. http://www.redhillstudios.com/#/projects/games/pdwii/. Accessed 26 March 2012
Rego P, Moreira P, Reis L (2010) Serious games for rehabilitation: a survey and a classification towards a taxonomy. 5th iberian conference on information systems and technologies (CISTI), pp 1–6, June 2010
Ma Roubidoux, Chapman CM, Piontek ME (2009) Development and evaluation of an interactive web-based breast imaging game for medical students. Acad Radiol 9:1169–1178
Sabri H, Cowan B, Kapralos B, Moussa F, Cristanchoi S, Dubrowski A (2010) Off-pump coronary artery bypass surgery procedure training meets serious games, IEEE international symposium on haptic audio-visual environments and games (HAVE), pp 1–5, 16,17 Oct. 2010
SAIC Inc (2012) OLIVE–On-Line Interactive Virtual Environment. http://www.saic.com/products/simulation/olive/. Accessed 7 March 2012
SAIC Inc, The Faculty of Medicine at Imperial College London (2008) Game-based learning for virtual patients–multi patients. http://www.youtube.com/watch?v=VhQ8MjdRq_4&feature=related. Accessed 7 March 2012

Scarle S, Dunwell I, Bashford -RT, Selmanovic E, Debattista K, Chalmers A, Powell J, Robertson W (2011) Complete motion control of a serious game against obesity in children. 3rd international conference on games and virtual worlds for serious applications (VS-GAMES2011), pp 178–179, 4–6 May 2011

Schonauer C, Pintaric T, Kaufmann H, Jansen -KS, Vollenbroek -HM (2011) Chronic pain rehabilitation with a serious game using multimodal input, International conference on virtual rehabilitation (ICVR2011), pp 1–8, 27–29 June 2011

ScienceApplications (2010) Olive: 3D hospital training. http://www.youtube.com/watch?v=MhzD0UO_nUY&feature=related. Accessed 7 March 2012

ScienceDaily 2009 (2009) Virtual patient helps train pharmacists of the future. http://www.danshope.com/news/showarticle.php?articleid=83. Accessed 7 March 2012

Skills2Learn, Ltd (2010) Skills2Learn virtual reality and 3D simulation examples. http://www.skills2learn.com/virtual-reality-case-studies.html. Accessed 6 March 2012

Sliney A, Murphy D (2008) Jdoc: a serious game for medical learning, 1st international conference on advances in computer-human interaction, pp 131–136, 10–15 Feb. 2008

Tate R, Haritatos J, Cole S (2009) Hopelab approach to re-mission. Int J Learn Media 1:29–35

The Partnership for Food Safety Education (2008) The food detectives fight BAC! game. http://www.fooddetectives.com/. Accessed 26 March 2012

University Medical Center Utrecht (2012) Air medic sky 1. http://www.airmedicsky1.org/. Accessed 26 March 2012

Vidani AC, Chittaro L, Carchietti E (2010) Assessing nurses' acceptance of a serious game for emergency medical services. 2nd international conference on games and virtual worlds for serious applications (VS-GAMES), pp 101–108, 25,26 March 2010

Vtnen A, Leikas J (2009) Human-centered design and exercise games. In: Kankaanranta M, Neittaanmki P (eds) Design and use of serious games, Springer Science + Business Media, pp 33–47

Waite M, Hawker S (2009) Oxford paperback dictionary and thesaurus. Oxford Paperbacks, Oxford University Press. http://books.google.es/books?id=8H5_od8I6pMC

Wang Q, Sourina O, Nguyen MK (2010) EEG-based serious games design for medical applications. International conference on cyberworlds (CW), pp 270–276, 20–22 Oct. 2010

Watters C, Oore S, Shepherd M, Abouzied A, Cox A, Kellar M, Kharrazi H, Liu F, Otley A (2006) Extending the use of games in health care. Proceedings of the 39th annual hawaii international conference on system sciences, HICSS '06. vol 5. pp 88b, 4–7 Jan. 2006

World Health Organization (2006) Constitution of the world health organization–basic documents (45th edn Suppl)

DDRGame (2012) Dance Dance Revolution for Wii, PS2, PS3, Xbox 360 and PC. http://www.ddrgame.com/. Accessed 29 March 2012

GF Engine

A Versatile Platform for Game Design and Development

Hock Soon Seah, Henry Johan, Chee Kwang Quah, Nicholas Mario Wardhana, Tze Yuen Lim and Darren Wee Sze Ong

Abstract The GF Engine is conceived to be a versatile platform for game design and development that is based on best-of-breed game components and technologies to facilitate related R&D activities as well as application and content development. The engine aims to provide a framework for the research community, from researchers, technologists, application developers, and content creators alike, to collaborate, share, and experiment with innovative ideas and creative endeavors. At the research end of the spectrum, researchers conducting R&D on specific components could utilize the GF Engine as an integrating environment for proof-of-concept validation. At the application end of the spectrum, content developers could turn to the GF Engine to create their applications and contents efficiently.

Keywords Game engine · Best-of-breed components · Future-proof · Abstraction and encapsulation · Rendering · Physics · Scripting

1 Introduction

The GF Engine, abbreviated here as GF, is a versatile gaming platform where the research community may use as a framework to experiment and innovate their research. At the same time, developers, game designers, and gamers can create or modify a game in an intuitive manner. GF does not reinvent the wheels but instead takes advantage of the state-of-the-art open source software packages. However, it

H. S. Seah (✉) · H. Johan · C. K. Quah · N. M. Wardhana · T. Y. Lim
School of Computer Engineering, Nanyang Technological University, Singapore, Singapore
e-mail: ashsseah@ntu.edu.sg

D. W. S. Ong
DSO National Laboratories, Singapore, Singapore

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_10,

will not be totally dependent on these packages. It builds wrappers around them and grows with their future additional features. Its modular architecture allows replacement of individual or all components it is built upon.

Up to now, making game requires intensive collaboration between game designers and game developers. By decreasing the involvement of developers in the pipeline, GF plays the role of the developers. This is a big challenge, especially when GF wants to achieve the same level of runtime performance provided by other specialized game engines. To support rapid game development, GF is divided into two layers of abstraction, the *core engine* for general purpose game programming and the *template-specific API* for rapid construction of a well-defined game genre.

The rest of this chapter is organized as follows. Section 2 discusses the objectives of GF, what is it conceived to be. This is followed by the design approaches behind GF in Sect. 3. In Sect. 4, the architecture of GF is discussed. Details of some of the components in the architecture are explained in Sect. 5. An application and coding example is given in Sect. 6, and we conclude in Sect. 7.

2 GF Engine Objectives

2.1 Evolving Platform

Studying the architecture of present game engines and simulation software, we can summarize that game engines consist of the following components which are widely available as standalone packages: graphics renderer, physics simulator, artificial intelligence module, networking, and asset database.

GF, as a game development platform, also contains the above-mentioned components. GF integrates and manages the interaction among these components in a way such that the game development process can be done as easily as possible. In addition, GF, aiming to stay future-proof, is designed based on a modular architecture. Various components in GF can be replaced, allowing GF to evolve and stay up to date.

Using abstraction and encapsulation techniques, class wrappers are implemented around each component to isolate them from the rest. These modules are integrated in a well-defined structure, which is applicable to many components currently available in both commercial and open source packages. The features of these components are manifested through the core of GF, i.e., the Sandbox Integrated Development Environment (IDE).

2.2 Platform for Everyone

Game programming is a complex combinatorial problem. Creating games, especially networked games, is very challenging. A typical decent game studio consists of many skillful programmers, artists, and game designers working together in a sophisticated work flow. Even experienced programmers or scientists from different fields find it difficult to create game or simulation for visualizing their experiment results. GF simplifies the game creation process. With its sandbox IDE, a user can easily realize his/her own game ideas without much programming efforts.

3 GF Engine Design Approaches

The currently available game platforms in the market fall into two categories: One that aims for the best game quality, in term of scale and performance. One that is very easy to use, with limited flexibility.

In order to achieve best performances, the former platforms focus on every single detail of the gaming hardware to achieve optimal resource utilization. They usually come with libraries that provide access to alter hardware details down to the bits level. For example, Killzone engine (Valient 2007) allows direct communication with the gaming hardware so that programmers can carefully design their codes, maximizing memory access speed, and number of cache hits. Unfortunately, these skills require advanced knowledge and steep learning curve.

On the other hand, the latter type of engines, which are targeting casual users, have their own downsides. For example, Little Big Planet (Media Molecule 2008) allows gamers to create their own games, but with a lot of restrictions. Users are only allowed to build their games using a set of predefined game logics. The only full flexibility that the users can enjoy is to organize their game objects freely in the game world.

GF aims to strike a balance between the two; emphasizing on the ease-of-use without much compromise on the game quality.

3.1 Simplified User Interface

GF provides a simple yet effective user interface (Fig. 1). It comprises three main components:

Game Area: The window where users load and play their games. In GF, the game is always running. Users can seamlessly add or remove game logics without the need of restarting the game.

Script Editor: Provides facilities for users to define their game logics.

Fig. 1 GF user interface

Object Inspector: Provides facilities to inspect and modify the current game state (including attributes that are not visually displayed in the game area).

GF maintains high integrity among the three components. Each is linked to the others. For example, every line of script that creates an actor will be linked with the corresponding actor in the Game Area. When users delete or modify that line, the corresponding actor will be updated accordingly. For every class defined in the script, there will be a button created in the Game Area to quickly populate the scene with objects of that class. When an object in the Game Area is selected, all its properties will be displayed in the Object Inspector. Users can view and even modify these properties. All changes made will be converted to script and inserted into the game editor.

With this interconnection between the three components of the interface, GF interface is highly streamlined and simplified. Three aspects of the game production are put into three parts of the interface: game definition, game testing, and game inspecting.

It is a challenge to design a simple yet intuitive Script Editor. To ease the scripting process, an editor needs to expose various features to its users. One common approach adopted by most editors is by means of buttons and drop-down menus. As a consequence, users may be overwhelmed with all these Graphical User Interface (GUI) elements. Instead, Script Editor adopts the principle of context sensitive content to maximize intuitiveness. Only functions that are available at the current context will be displayed. In addition to minimizing the number of available functions, the number of contexts that can be switched from the current one are maximized. At any one time, users are expected to work with a

block of game definition in the Script Editor. All possible options for that block should be neatly displayed. The most appropriate location to prompt these options is at the text cursor, where they can be immediately applied. This design decision leads to two benefits:

First, text cursor is an intuitive indicator for users to make changes to the current block of game definition.

Second, in traditional approaches, dialog boxes will distract the users to certain degree. In GF, selectable options are both displayed and applied at the cursor location, keeping users' attention in one place, thus promoting faster scripting.

3.2 Simplified Work Flow

The contemporary work flow to create a game is complicated. The game ideas come from game designers and the artworks come from the artists. These requirements are communicated to programmers to produce the actual game. It is common to have miscommunication in the process, resulting in constant changes and tweaks being made until the game is finished. Furthermore, current game engines have to compile game codes written in a programming language to machine language before execution. The game needs to be restarted each time tweaks and changes are made, resulting in significant amount of time being wasted. These include the time needed to start the game (usually longer in the debug build), and the time to bring the game state to where it can show the changes (i.e., if the new change involves more game characters, the tester needs to recreate them to see the change). Also it would be useful for game designers to try out and see their ideas instantly. To simplify the work flow, GF reduces:

Communication time: GF enables game designers to directly create games. GF Factory emulates the role of programmers and translates game ideas to executable codes. This is achieved through the user interface and Rule-Oriented Simulation Architecture Language (ROSAL) language, which is based on Aspect-Oriented Programming (AOP). AOP itself is an extension of Object-Oriented Programming (OOP), see Sect. 5.3.

Time needed to verify the game changes: GF implements a feature, namely Live Sandbox, which enables the game to stay running and changes are instantly applied without the need of restarting the game. As a result, GF becomes a WYSIWYG (what you see is what you get) game editor. This also helps maintaining the game designers' stream of thought flow.

4 GF Architecture

This section describes the GF architecture. It also discusses the design decisions made to overcome challenges encountered and achieve the project objectives.

4.1 Components and Abstraction Layers Isolation

GF has a modular architecture (Fig. 2) allowing the replacement of one or all components it is built upon. In order to do this, GF needs to isolate the components it uses while still be able to extract the component's functionality. Currently, GF has three isolated components: rendering engine OGRE (The OGRE Team 2012), physics simulation engine NVidia PhysX (NVidia 2012b) and input engine OIS (Wrecked Games 2012). Furthermore, GF has two layers of abstraction, core engine layer and template-specific layer which are isolated as well. This section discusses the design to achieve these requirements.

4.2 Unified Data Structure

The first challenge comes from the different data representations of the components for the same thing. For example, the rendering engine uses (*x, y ,z, w*) for Quaternion, while the physics engine uses (*w, x, y, z*). These low-level data are

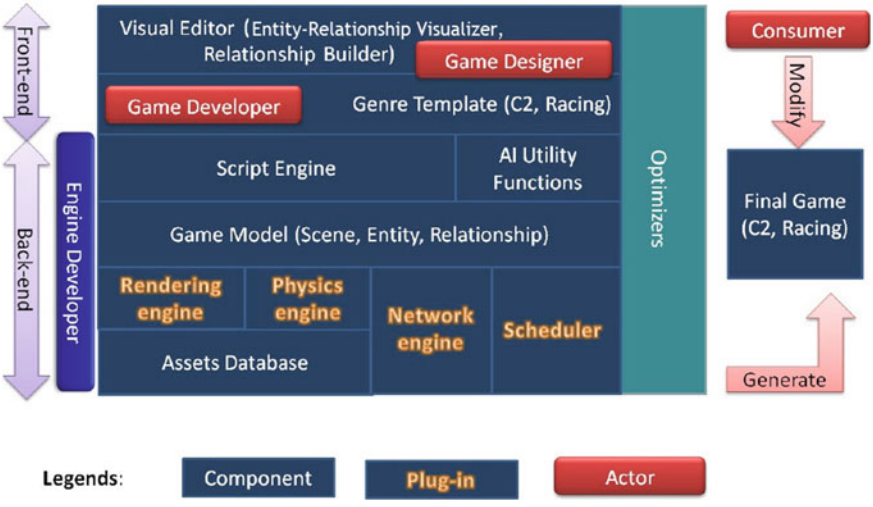


Fig. 2 GF architecture

exchanged between components at a very high rate for every frame and thus an efficient method is essential for data conversion and exchange.

Object transformation is usually represented by a 4×4 matrix for translation, rotation, scaling, and some tweaking like skewing. After some analysis, the matrix representation is inappropriate due to the following reasons:

The physics representation of object neither uses scaling nor skewing, only translation and rotation are considered. It is logical since object's size and shape are not changed in a physical environment. In some cases, such as skinned objects or soft-body objects, those changes are represented at a lower level.

The matrix size is too large for low-level data exchange, which happens at very high rate every frame. We use only translation and rotation using quaternion, i.e., two 4D vectors, which also save half of the exchange bandwidth.

In many cases, the matrix representation is decomposed back to translation vector and quaternion for computation. Thus, involving matrix will add up computation cycles to the data exchange.

Due to the above reasons, we pick the translation vector and quaternion as the basic data types between the rendering and physics engines. The next step is to solve the data type representation between them. Here are several approaches:

- Write conversion macros between the two corresponding types. For example, Vector3_to_NxVec3 and NxVec3_to_Vector3. This is the most efficient way in terms of performance, but is harder to comprehend for developers.
- Write an intermediate type which has automatic conversion operators to corresponding types. For example, we write class Vec3 which can be used in place of both Vector3 and NxVec3. This is the most elegant way for programmer while introducing much more CPU cycles for conversion.

We balance between the two approaches by introducing an intermediate class, derived from one engine, which has automatic conversion to the other type. Using the second approach saves half of the bandwidth. In the case of changing engines, the two intermediate classes (Vec3 and GFQuad) need to be rewritten.

4.3 Conflict of Data Organization

Different components have different ways to organize their data structure. The OGRE rendering engine uses local coordinate system to organize objects while the PhysX engine uses global coordinates. Furthermore, each object in OGRE must be given a unique name, which is quite inconvenient to instantiate an object. To resolve the coordinate conflict, several things are need to be considered:

- The direction of data exchange.
- The coordinate system we use.
- How to maintain features in hierarchical coordinate system.

4.4 Resolving Runtime Data Type Checking

In order to trap physics events, such as collision or ray cast hit, the physics engine provides a pointer to store *userdata* to identify which object is responsible for the event. It is not possible to get the type information of these objects from the event callback without carefully designing the class structure. Since GF uses C++, which does not store dynamic type information like other advanced languages, it can only store one single type in the *userdata* of the physical objects. This type will be responsible for implementing the event handlers. Polymorphism will play the role of type identification. There are multiple options for this type selection:

Store the Physical Object. This approach hides the physics component interface although we need to store a pointer to GameEntity, since the entity is what we actually need.

Store the GameEntity object. This approach returns directly the object we need but expose dependencies on the physics component.

Identify the general physical events and make wrappers around them. GameEntity will adopt this wrapper as one of its parents. This is GF's choice since it has all the benefits of the other approaches.

Another issue is to support type checking at the script level, because ultimately, most of the physical events need to be accessible in the script space.

4.5 Template API Isolation from Core Engine

GF divides object declaration into two levels. At the core engine level, objects are defined with general purpose properties, while at the template level, objects are added with more specialized properties. For example, the C2 (Command and Control) Template (see Sect. 4.7) needs to add other types such as Selectable Object, Movable Object, Attacking Object that can be mixed with the core object types in a variety of ways.

In GF class design, we adopt multiple inheritances. Functionalities are implemented as classes, such as PhysicalEventHandler that is actually an interface with all pure methods, or ScriptID, which is a constructor wrapper to record the object address for object identity at script space. Classes like ScriptID are treated as self-implemented interfaces. Other classes, which want to possess these features, can be extended from these base classes, along with their own primary parent classes. Multiple inheritance enables GF to create combinations of features. This is especially useful at the template level, where many classes share a subset of features. For example, Tank is extended from AcquirableObject and SelectableObject; whereas Soldier is extended from SelectableObject and MovableObject, etc.

GF supports development of multiple types of game by providing a template to each game genre. The template is developed by a programmer and exposing facilities that a game designer can use for faster game development.

4.6 Core Engine Class Design

The class relationship for the core engine is illustrated in Fig. 3. The blue outlined classes are abstract classes. The red generalization lines are virtual inheritance.

The core engine classes are designed with the following wrapping principles:

Step 1 Identify main objects from the underlying engines. For example, the physics engine consists of rigid object, composite object, controllable object, vehicle objects, etc.

Step 2 Wrap these objects so that the final interface contains no dependency on these objects or the relationship between these objects

Step 3 Using the wrapped interface, connect objects from different engines together to form a final object which is visible at scripting level

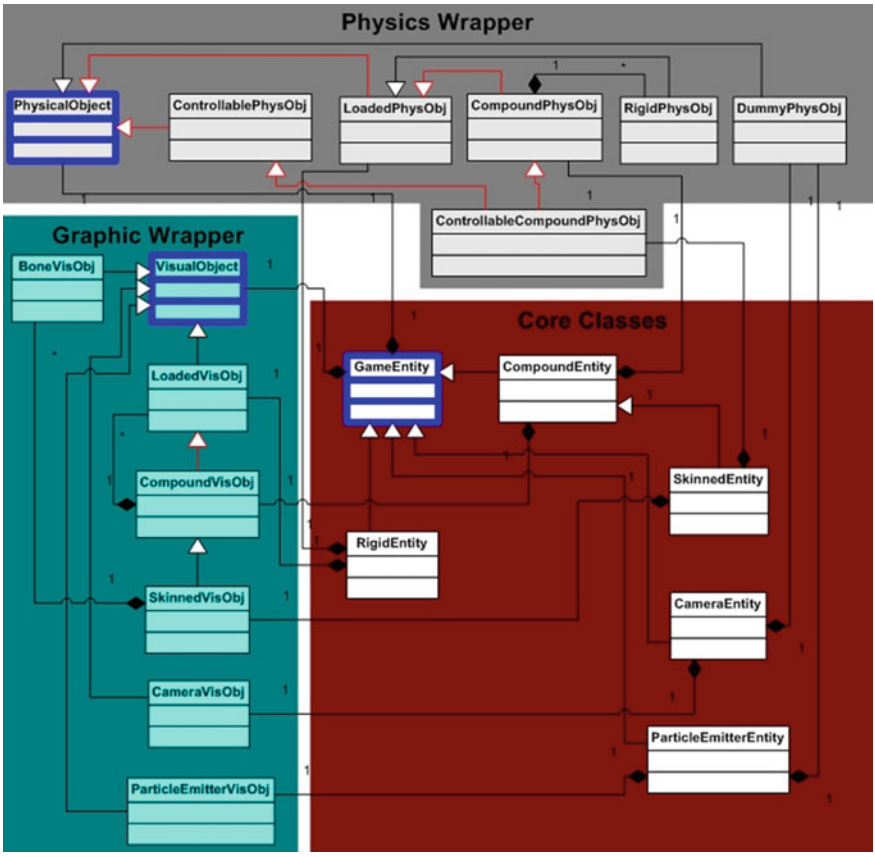


Fig. 3 Core engine class diagram

In Fig. 3, we identify the commonly used objects in Graphic Wrapper and Physics Wrapper. These objects can be found in other similar engines as well. Therefore, when a better engine is introduced, GF can simply adopt it by writing a new wrapper based on the existing interface.

4.7 Template Class Design: C2 Template

The logic within a template should derive only from the core classes to be independent of the underlying components. Furthermore, the class design should be flexible for a complex set of object with mixed behaviors. The current design is illustrated as in Fig. 3. The blue outlined classes are abstract classes. The cyan classes are self-implemented interfaces described in the previous section.

Every object in the C2 Template is derived from a single core class and a combination of self-implemented interfaces. One object can inherit features from these self-implemented interfaces by simply extending from them. Some interfaces work with their container, for instance, the SelectableObject is manipulated by the EntitySelector. These classes reflect our vision of a C2 game, where objects can be selected and commands can be issued to each object.

5 GF Engine Components

Figure 4 shows the various components in the architecture of GF. Due to space constraint, only the details of the rendering, physics, and script components are discussed in this section. For asset database, instead of using proprietary data format or tool set, GF adopts the emerging standard COLLADA (Khronos Group 2012) for all kind of digital assets. This allows the use of common digital content creation tools such as 3ds Max and Maya. Game artists can then make the best use of their experience with their favorite software without having to learn new ones. An example of artificial intelligence component can be found in (Wardhana et al. 2009).

5.1 Rendering Component

Besides the usual Forward Shading, GF rendering system also supports the Deferred Shading (Hargreaves 2004) technique. GF currently uses OGRE's features for rendering, and shading languages supporting Cg (NVidia 2012a), GLSL and HLSL.

Material Rendering GF supports material down to the shader level. The shader is written by graphics programmers and can be reused by artists.

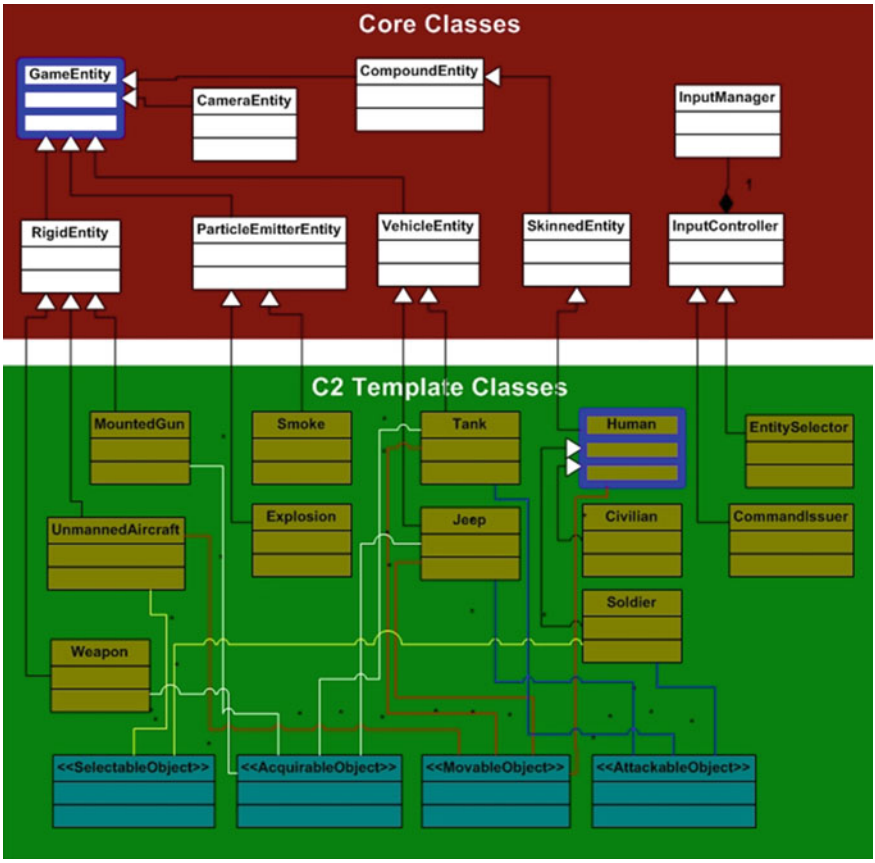


Fig. 4 C2 template class diagram

Compositor System GF has a complete system for full screen effect using OGRE compositor framework. These effects are achieved by applying shader materials to the whole screen, or render target texture. GF supports multiple rendering windows, which are incorporated into its GUI. Users can add more cameras and display these cameras by using these rendering windows.

Debug Renderer A debug renderer is supported for all physical bounding shapes, axes, and constraints. Objects can be switched to wire-frame or dot mode to clearly see hidden details (Fig. 5).

Deferred Shading Deferred Shading is a modern rendering technique leveraging the new rendering hardware platforms. It becomes more and more popular in the recent years. Many award winning games (e.g., Killzone 2, God of War 3, Little Big Planet, etc.) adopted the Deferred Shading technique and successfully created the best real-time graphical presentation up to now.

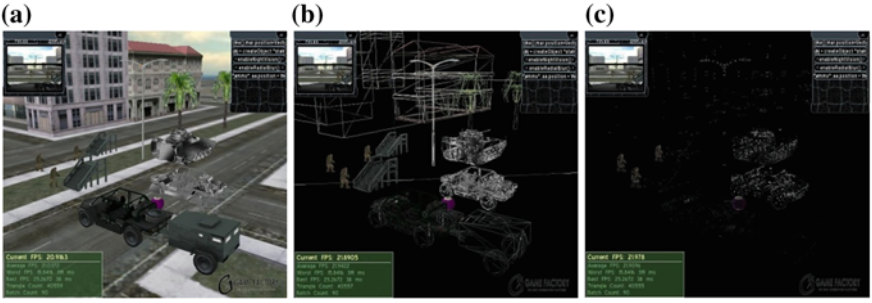


Fig. 5 Debug renderer **a** Fully rendered mode **b** Wire frame mode **c** Dot mode

The strongest aspect of Deferred Shading is the low cost of rendering a light. A traditional rendering engine usually limits the number of lights that simultaneously affect an object. The limitation ranges from 4 to 8. If this limitation is exceeded, one or more lights will be disabled. In the case of dynamic lighting, going in and out of the limitation may result in a blinking scene. This usually leads to a strict lighting design that limits the number of dynamic lights in the entire scene to 4 or 8. By introducing low cost lighting, Deferred Shading provides flexibility in lighting design for the scenery and even game play.

Furthermore, by shifting lighting computation to the end of the rendering passes, shadow manipulation becomes much easier. Shadow casting lights can now share a common shadow map texture, which is not possible in traditional rendering engine. This eliminates memory constraint imposed in shadow casting lights, allowing a larger number of shadow casting lights in the scene.

With the availability of depth surface in Deferred Shading, light can be modified to be used in various circumstances such as decal or caustic. Grouping objects into different groups, z-sorting can be applied to these forms of lights. For example, static decal will be applied only on static objects and dynamic objects can obscure it; whereas dynamic decal will cast on all objects.

Special Effects Deferred Shading renders all objects into three flat surfaces: normal surface, depth surface and color surface. These surfaces provide rich information for a lot of effects, such as Depth of Field and Screen Space Ambient Occlusion (Fig. 6). Furthermore, the cost of such effects is reduced significantly as each pixel is computed only once for the effects.

5.2 Physics Component

The current physics simulation in GF is NVidia PhysX. This is a close source engine with license to use freely for commercial purposes. So far, this is the only engine with hardware acceleration, supported by NVidia, the flagship company that produces graphics cards, where some of them have PhysX hardware support.

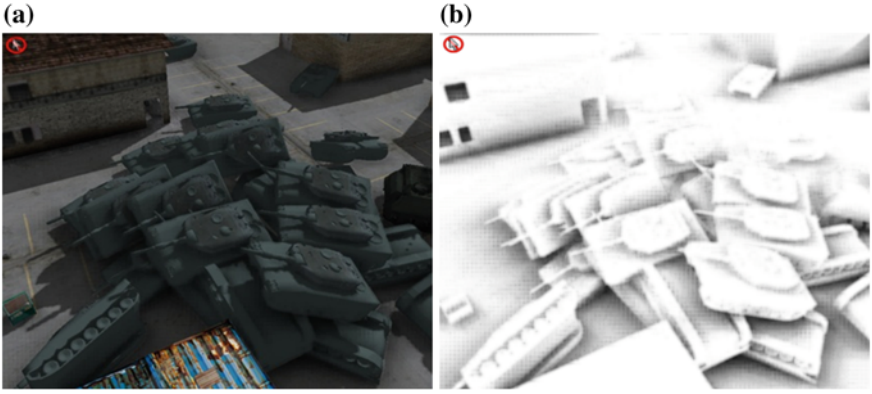


Fig. 6 Screen space ambient occlusion **a** Fully rendered frame **b** Screen space ambient occlusion component

Rigid Body Simulation and Collision Detection Rigid body is an object that keeps its size and shape unchanged after the collision, i.e., an unbreakable object. Rigid Body simulation is performed extremely fast in PhysX with Broad Phase.

Collision Detection and multicore support. GF supports loading complex data structure to represent the Rigid Body shape. PhysX supports from basic shapes with low cost in collision detection (such as box, sphere, capsule, cylinder, etc.) to advanced structures (such as height field, convex mesh, etc.). GF supports ray casting for multiple uses. Ray Casting is a collision detection manner using a ray casted from an origin to infinity. This is performed really fast and allows a large numbers of rays in one frame. Ray Casting is used in mouse pickup and way-point generation.

Physical Constraint Physical constraint links objects together in a specific way. GF supports multiple types of physical constraints, from hinge joint, sphere joint, sliding joint to 6 degrees-of-freedom (DOF) joint. One of the applications of physical constraint is Rag Doll physics. GF has implemented high quality Rag Doll physics for all skinned objects (human, animal or other biological entities). This information of Rag Doll can be loaded from files created with digital content creation tools, such as 3ds Max. The visual and physical information is linked using name matching (in this case, the bounding shapes have the same names with the corresponding bones).

GF also supports breakable constraints. The constraint can be given a limit on the force that it can sustain. If the force exceeds this limit, the two connected physical objects will be disjointed. This is very useful to model destructible objects, such as house and car.

Controllable Object GF uses a physics simulation for all objects inside the game. Therefore, it is hard to directly control an object movement using only physical input, such as force and torque. These are the reasons:

It is hard to make object stand still. We can prevent the sliding of the object by setting a high friction value. However, this approach may make the object hard to

be moved, or move inconsistently on different surfaces. It is hard to make object move without rotating since any force deviated from object's center of mass will result in a rotation. It is hard to make object climb up small obstacles. It is hard to make object continue its movement when collides with obstacles.

GF also has designed a controllable class that solves all the problems mentioned above. We use a capsule shape for the object and handle all collision events manually. Furthermore, GF allows switching from controllable state to any other physical states. For example, the soldier can turn from controllable physical state into compound physical state (for Rag Doll death animation).

5.3 Script Component

GF exposes its features to game designers through a scripting language. A set of selectable sections of the script can be exported to the end-users (i.e., gamers) for further modification. GF has chosen script, instead of pure GUI approach, for more flexibility and scalability. In order to meet the requirements of GF, the script subsystem needs to provide:

A supporting IDE with easy-to-use interface for rapid game development.

A scripting language, instead of pure GUI approach, for finer adjustment. The language should conceal programming-specific details with its advanced features.

Runtime performance of the final game should not be compromised.

Support classic programming paradigms, such as Object-Oriented or Procedural, to be friendlier to programmers.

ROSAL Design Principles Previous game engines are targeting programmers as their users. Therefore, the integrated languages of these engines usually focus on Object-Oriented Programming (OOP) design, which is very familiar to programmers and allow construction of complex game structures. Although being the most popular programming paradigm at the moment, an OOP language suffers from high maintenance cost due to cross-cutting functionalities which span over the boundaries of class encapsulation (Constantinides and Skotiniotis 2002). This problem is solved in Aspect-Oriented Programming (AOP) (Kiczales et al. 1997) language, which is an extension of OOP and separating cross-cutting functionalities into individual concerns. However, AOP languages are not popular due to its complexity.

By changing the target users to game designers and gamers, GF scripting language needs a different approach. ROSAL is designed around the concept of AOP which has enough modularity for simulation application while still being an easy to use language.

ROSAL has a number of features supporting the reading and maintenance of the code. First, ROSAL has a syntax resembling human language. Second, its IDE can display explanatory information for every piece of the code. Third, the IDE Diagram Generator can help to illustrate complex architecture and relationship between game entities. Finally, the IDE Filter provides quick information retrieval and code organization.

ROSAL, as its name suggests, focuses on the construction of rules. The programmer just declares the rule in a way that is very similar to its real-life counterpart. The language takes care of how the rules are implemented.

ROSAL is a multiparadigm programming language. It is based on OOP and thus supports inheritance, polymorphism, and encapsulation. Furthermore, it has separation of cross-cutting concerns borrowed from AOP. This helps in organizing the program structure. It is built upon Lua (Ierusalimschy et al. 2007) and Metalua (Fleutot 2007). Lua is an effective scripting language with high customization and its semantics can be changed using well-defined meta-methods, while Metalua is a Lua parser generator written in Lua. In addition, ROSAL also encompasses a C++ Luabind (Wallin and Norvig 2003) mechanism in which C++ objects could be extended via a scripting manner, which allows rule-based scripting functionalities to act on the objects. With ROSAL in-conjunction with C++, we aim to provide developers better choices as compared to the more general purpose programming languages Actionscript (Crawford and Boese 2006), C#, Python, Java, etc. This coding choice enables more compute intensive and hardware-related tasks to be carried out in C++ while allowing the rules to be scripted in ROSAL in a more intuitive manner.

The language supports both imperative and declarative programming. When describing a rule, a programmer just declares which objects are involved by defining a condition. The language will determine how to retrieve these objects in a CPU-saving manner. After the objects are identified, a sequence of instructions is carried out. By doing this, the language can work totally at class level, which means the programmer is relieved from the burden of effectively storing and retrieving objects.

6 Application and Coding Example

Through a typical game production pipeline, we illustrate the use of GF to develop a 3D computer game. GF acts as a reusable library that provides objects and functionalities, as well as adding new functionalities and gaming components.

6.1 Game Production Pipeline

Figure 7 shows a typical game production pipeline. The production process consists of designing game resources (objects and characters, their animation), and then importing them into GF before performing the necessary programming and scripting to implement the gameplay with respect to the game story.

Design of Game Resources The game resources consist substantially of (but not limited to) the game objects, characters, their animations and the necessary texturing materials. These 3D characters, comprising their geometry, physical attributes, animation, and materials, can be designed and created by common practices such as

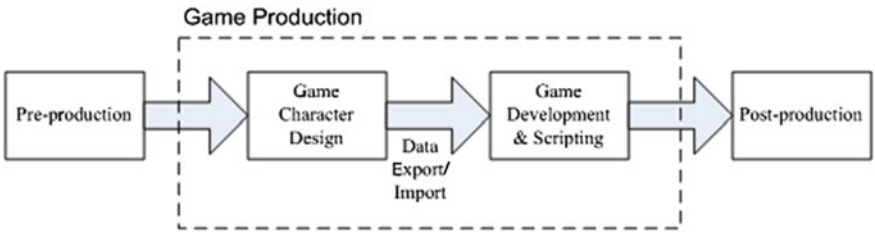


Fig. 7 Typical game production pipeline

via using of software tools like 3ds Max. These resources are exported into compatible formats that will be loaded into the game engine.

Types of Game Objects and Characters There are several basic types of objects and characters in GF. We defined them as entities: static rigid entity, movable rigid entity, skinned entity, and interactive entity. The characteristics of these entities are decided during the design and creation phase. These entities are then inherited so that their fundamental features are further extended. For example, an animated soldier character is inherited from a skinned entity to allow the surface mesh model to deform according to its underlying skeleton joints.

Programming and Scripting Coding in GF using C++ allows extensions to the general functionalities, objects, graphics and physics algorithms, and be exposed in a form that is callable by the rule-based ROSAL scripting that allows non-expert developers to simply code real-time discrete game elements and events.

6.2 First-Person Shooting Game Example: C2 Game

This is a first-person shooting military-based game, which makes use of the graphics, physics and audio components of GF, as well as extending the engine's core functionalities. Scene and object creation are done through 3ds Max (Fig. 8) with NVidia PhysX plug-in that enables the handling of dynamics and collisions.

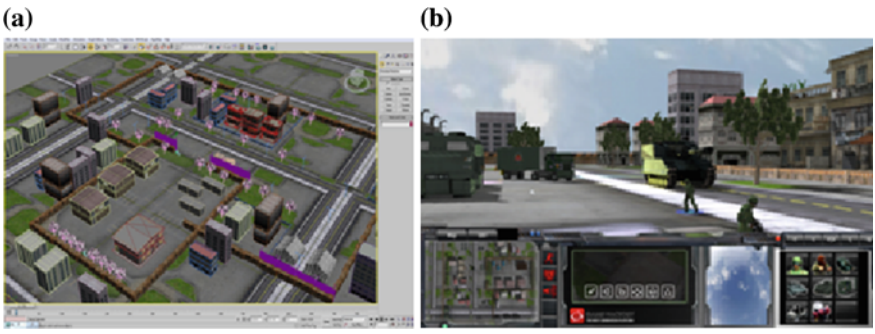


Fig. 8 First-person shooting game **a** Scene created in 3ds Max **b** Scene imported into GF

The fundamental object entities of GF are extended into more specific entities, such as the soldier and vehicles, coded in C++ for computing effectiveness:

```
// inherit Soldier from SkinnedEntity
class Soldier : public SkinnedEntity {
   public:
      virtual void FrameMove();
      // extending the each frame event
      ...
      void TakeCover();
      // new functionalities for Soldier
      void AttackEnemy();
      ...
};

// inherit Vehicle from MovableRigidEntity
class Vehicle : public MovableRigidEntity {
   public:
      virtual void FrameMove();
      // extending the each frame event
      ...
      void Steer();
      // new functionalities for Vehicle
      void Brake();
      ...
};
```

The main step for creating the game-play is loading of scene and characters via scripting:

```
-- to load scene
Scene.loadScene('MilitaryPrototype')

-- create extension soldier class in ROSAL
inherit Marine from Soldier

-- setting attribute of a marine
Marine.HealthPoint = 100
Marine.WhichTeam = 'A'
Marine.Ammo = 10000
...

-- create an instant of Marine type
mm[1] = Marine()
...

-- extending other gaming entities/characters
inherit Jeep from Vehicle
inherit Tank from Vehicle
...
```

Controlling the game entities

```
if key == 'W' then
  -- animation from the resource in response to key 'W'
    _getBaseObj(Marine):animation('walk_forward')

if key == 'S' then
  -- animation from the resource in response to key 'S'
   _getBaseObj(Marine):animation('walk_backward')
...
```

Rules that operate among the game entities

```
rule 'marine attack enemy'
    actor marine as Marine, enemy as Marine
    when distance(marine, enemy) < 30
       and marine.HealthPoint > 0.5
          marine.OpenFire()
    end
end

rule 'marine flee'
    actor marine as Marine, enemy as Marine
    when distance(marine, enemy) < 30
       and marine.HealthPoint > 0.5
          marine.Flee()
    end
end
```

7 Conclusion

Many software development paradigms and methodologies serve a common purpose: to address the myriad forms of inefficiency in the software development process and the brainchild of any such methodology originates typically from practical experience. However, the solution of adopting all practices of a methodology may be difficult in the case of game development.

To date, many game engines abstract production artifacts as data (content), behavior/logic and the runtime (engine). Redeployment of software for a similar game genre is superficially achieved through either modification of data layer or "rewiring" of the code and this typically requires the user to have sound knowledge of the game engine's technology. One of the key emphases of GF's design is reusability which is achieved through custom application templates. Such a design brings forth the advantage of accelerating the generation of common features while isolating the user from the low level implementation details. Under the circumstance when development is necessary for a specialized need, such effort can be "amortized" over future reuse scenarios through templates which are more efficient and manageable than raw code manipulation.

Another innovation of GF's overall design is its ability to stay "future-proof". Through software frameworks, only high level adaptors need to be built for

specific functions that pertain to a certain genre of application. The implementation details will be totally abstracted from the user such that as the technology of a particular field advances, the system can remain updated by integrating the best-of-breed components into it. This approach contrasts greatly from the monolithic design of many proprietary game engines whereby the integration of auxiliary components may be difficult.

While the GF framework is functional and operational, many desirable features such as automated terrain generation, dynamic generation of destroyable objects, advanced artificial intelligence for non-player character, networking features to support MMOG, and expanding GF to run on diverse platforms are missing. These are in our future R&D plan.

Acknowledgments We would like to acknowledge the support by the Singapore Defence Science and Technology Agency under Project Sigma and contributions of Gabriyel Wong, Hai Nam Pham and Yin Mun Wong during the early development of GF, which was known as Game Factory then.

References

Constantinides C, Skotiniotis T (2002) Reasoning about a classification of cross-cutting concerns in object-oriented systems. In: Costanza P, Kniesel G, Mehner K, Pulvermüller E, Speck A (eds) Second workshop on aspect-oriented software development of the German information society

Crawford S, Boese E (2006) Actionscript: A Gentle Introduction to Programming. J Comput Sci Coll 21(3):156–168

Fleutot F (2007) Metalua manual. http://metalua.luaforge.net/metalua-manual.html. Accessed 12 June 2012

Hargreaves S (2004) Deferred shading. Game developers conference. http://ambility.com/DeferredShading.pdf. Accessed 12 June 2012

Ierusalimschy R, de Figueiredo LH, Celes W (2007) The evolution of Lua. In: Proceedings of the third ACM SIGPLAN conference on history of programming languages, pp 2-1–2-26.

Khronos Group (2012) COLLADA - 3D Asset exchange schema. http://www.khronos.org/collada/. Accessed 12 June 2012

Kiczales G, Lamping J, Mendhekar A, Maeda C, Lopes C, Loingtier JM, Irwin J (1997) Aspect-oriented programming. In: Aksit M, Matsuoka S (eds) ECOOP'97 object-oriented programming, lecture notes in computer science, vol 1241. Springer, Berlin, pp 220–242

Media Molecule (2008) Little big planet. http://www.littlebigplanet.com/. Accessed 12 June 2012

NVidia (2012a) Cg Tutorial. http://developer.nvidia.com/node/76. Accessed 12 June 2012

NVidia (2012b) PhysX. http://www.geforce.com/hardware/technology/physx. Accessed 12 June 2012

The OGRE team (2012) OGRE. http://www.ogre3d.org/. Accessed 12 June 2012

Valient M (2007) Deferred rendering in Killzone 2. http://www.guerrilla-games.com/publications/dr_kz2_rsx_dev07.pdf. Accessed 12 June 2012

Wallin D, Norvig A (2003) Luabind. http://www.rasterbar.com/products/luabind/docs.html. Accessed 12 June 2012

Wrecked games (2012) Object oriented input system. http://www.wreckedgames.com/forum/index.php/board,6.0.html. Accessed 12 June 2012

Wardhana NM, Johan H, Loh PKK, Seah HS, Ong DWS (2009) An efficient connection graph for waypoints in virtual environments. In: Proceedings of international conference on computer games, multimedia and allied technology (CGAT), 273–282

A Virtual CNC Training System

Peiling Liu and Cheng-Feng Zhu

Abstract Simulation of machining process for CNC training is significant, given its lower cost and risk-free nature. The drastic decrease of the cost of computing products, coupled with the worldwide price increases in material and machine tools means that virtual CNC training using computerized modeling and simulation is a cost-effective and sustainable approach to technical and professional education in manufacturing applications. The virtual CNC training system is developed for simulation of multiple machining processes. It is particularly important in the training of knowledge-intensive high speed and ultra precision machining. The in-process model proposed for the virtual CNC training system is presented in detail, followed by case studies of the industry training applications. Compared with conventional on-site manual training or e-learning, the virtual CNC training system greatly increases learning efficiency and effectiveness of trainees, and improves cost saving in terms of machine and material uses.

Keywords Modeling and simulation · In-process model · CNC machining · Virtual training

1 Introduction

As one of the major clusters of Singapore manufacturing industry, precision engineering contributes about 6 % Singapore's GDP. In addition, it is pivotal to the competitiveness of the whole manufacturing sector by providing support to electronics, medical, healthcare, and automotive industries. According to Singapore Precision Engineering and Tooling Association (SPETA), the output of

P. Liu (✉) · C.-F. Zhu
Singapore Institute of Manufacturing Technology, 71 Nanyang Drive,
Singapore 638075, Singapore
e-mail: plliu@simtech.a-star.edu.sg

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_11,

precision engineering in Singapore accounted for more than a tenth of Singapore's total manufacturing output. As a cutting edge technology, CNC machining produces essential inputs for virtually all types of manufacturing products for different applications, including injection mold, sheet metal die, casting die, jigs and fixtures, and other special tools. CNC technology has been widely used in Computer-Aided Manufacturing (CAM), High Speed Machining (HSM), Ultra Speed Machining (UPM) etcetera. The extensive use of CNC significantly improves the productivity of precision engineering but has caused a shortage of skilled technicians or machinists, especially in the knowledge intensive areas such as HSM and UPM. Training of skilled machinists is therefore a crucial yet challenging job. A qualified HSM machinist should have good knowledge of machining, understand the operation of the machine tools, and be able to do planning for machining process. Traditionally, trainees acquire their operating skills in several years through observation and reference to the operation manual. After which, they would learn to operate machines for themselves under the guidance of experienced operators. The acquisition and maintenance of real CNC machines, the consumptions of real materials in machining, and the setup and maintenance for workshops, all contribute substantially to the high cost of conventional CNC training. Cost-effective and safe CNC training is thus highly desired. While CNC training using real CNC machines is necessary, the use of Virtual Reality (VR) technology to support CNC training has been a popular topic recent years.

2 Current Status of Virtual CNC Training

Comparing with CNC simulation applications which are expensive and mature, Virtual CNC Training System is still primitive. The NC simulator developers are not very active to provide training system because the training software market is logically smaller than production software. More important, NC simulator developers need to revamp graphics engine or geometry kernel to suit education game use. None of the leading NC simulators has any CNC training capability. This leaves the development of Virtual CNC Training System to machine tool vendors and schools who do not have expertise to develop a good graphics engine.

The CNC control vendors developed their own training system. For example, Siemens developed SinuTrain, which is CNC training software. It runs on PC and suitable for training purposes and self-study as it is for writing programs and simulation. It serves for writing and simulating NC programs on a PC, based on the DIN 66025 programming language as well as the products ShopMill, ShopTurn, and ManualTurn+ and language commands for SINUMERIK® 810D, 840D, and 840Di controls, all are Siemens products. Programs written with this software can be used on real machines. A prerequisite is that the SinuTrain software is adapted to the SINUMERIK control on which the program is to be executed. This adaptation must be carried out by specially qualified personnel, e.g., from Siemens. It is

important to stick to Siemens and the machine tool manufacturer's instructions when adapting the software. No liability is accepted by Siemens if these requirements are not adhered.

We evaluated several vender specific Virtual CNC Training System. They have very good GUI which has been customized to vender's CNC controller, some even have touch tablet that simulates operation panel. However, the cutting simulation is rough and primitive, despite the sound and chip flying animation.

The chapter (Avgoustinov 2000) discusses the possibility to use VRML for scientific, educational, and even industrial purposes. As an illustration, it showed how VRML can be used as inexpensive means for simulation of one of the most interesting but also most time and resource consuming areas of the CAM—machining of complex parts.

There are many research and publications on Virtual CNC Training in last 10 years. Most of the published graphics engines are based on VRML and Java 3D. They explored internet-based CNC training and remote NC simulation, etc., which are futuristic but not practical at this moment. The computer hardware is very cheap now; there is no need to run a training system over internet. The remote graphics over internet is not necessary (Qiu et al. 2001; Ong and Mannon 2004). Some of them use flash movies such as micro media to do animation, which could only be used to pre-fixed scene.

In VRML, the realization of dynamic material removal during a machining process remains a problem. Some commercial software such as Deneb's virtual NC can export a VRML animation to describe a machining process. Nevertheless, during the cutting process, the geometry of a workpiece remains unchanged. The reason is that VRML does not support set operations among geometric objects such as union, intersection, and difference. This makes it difficult to simulate the change in geometry of a workpiece under cutting.

In layman's term, the current virtual CNC training systems are educational games that lack the realistic feeling of machining, which is quite different from realistic simulators, such as the flying simulators that are used to train pilot.

The next generation virtual CNC training is to provide knowledge-intensive CNC training, for the future skilled machinist of precision engineering (PE) industry through pervasive physics-geometrical modeling and simulation of multiple machining processes, especially high speed and ultra precision machining. To realize this vision, we need a new in-process model (IPM) that is deformable and precise.

3 Fundamentals of In-Process Model

The IPM represents the state of the product at each step in the machining process. Geometrically constructed in 3D, IPM can be used to reflect the results of the machining operations. This model allows the user to visually verify the accuracy of machining operations and the correctness of machining sequence. It can be automatically re-generated when there are changes in the product design,

machining parameters, or sequence of the operations (IMTI 2002). In this section, the techniques for geometric representation of IPMs for traditional manufacturing simulation are presented.

3.1 Boundary Representation

Naturally, boundary representation (B-reps) is the first choice for IPM representation as it is widely used in geometry modeling with commercial Computer-Aided Design (CAD) systems. With the aid of B-reps, applications can be developed through CAD development tool kits which are usually based on their CAD kernel modelers. Sharing a common geometry model with CAD, IPMs can facilitate seamless integration of CAD/CAM, and Computer-Aided Process Planning (CAPP). Liu (1991, 1992) worked on design automatic forging and manufacture system for pre-form forging using B-reps IPMs. Park et al. (2003) reported a prismatic B-rep IPM generation method that employed a polygon extrusion algorithm to sweep a ball-nose cutter. Even though great efforts have been made to improve the robustness of B-rep-based Boolean operation, the application using B-rep IPM is often limited to 2.5-axis milling (Fig. 1).

3.2 Sectional Representation

Since the integrated B-rep IPM cannot be created inside a CAD geometry model, an ad-hoc cross-section and wire-frame-based approach was proposed in (Liu 1991, 1992). The aim was to use a series of parallel cross-section drawing to represent 3D shapes. Figure 2b depicts the sectional representation of circled 3D shapes in Fig. 2a. Based on the same cross-sectional concept, I-DEAS from SDRC used water level cross-section as an IPM for generative machining. In a traditional NC programming environment, a significant amount of time would be spent for the

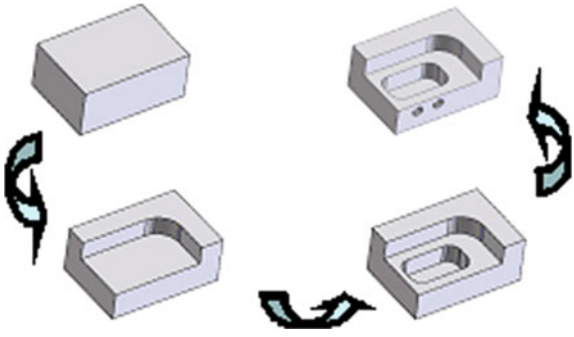


Fig. 1 2.5-axis IPMs

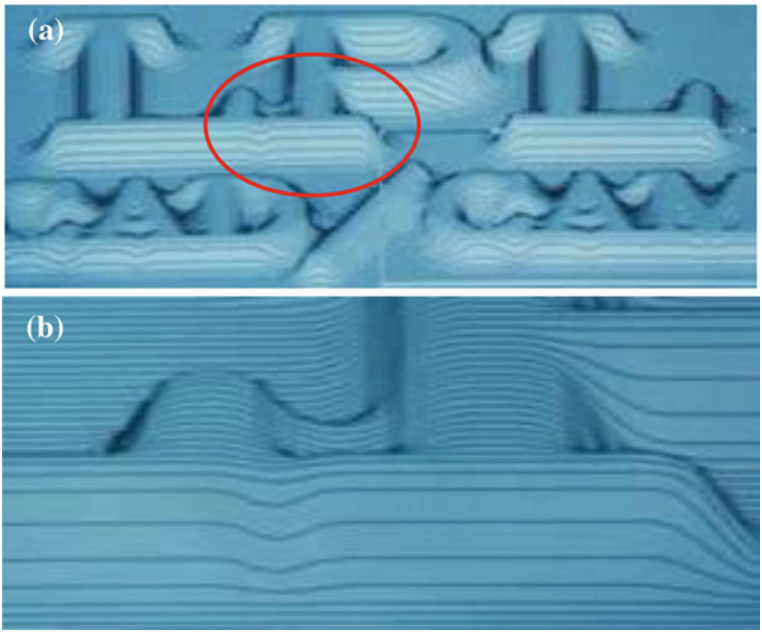


Fig. 2 Sectional representation. **a** 3D shape. **b** Sectional representation of 3D shape

purpose to visualize the in-process stock at different stages of machining process. With I-DEAS, the wire-frame section in-process stock model can be created for downstream applications such as tool-path generation, process planning, fixture designing, and clamp positioning.

A part can be sectioned along the X, Y, and Z axes (Fig. 3). The Z-axis section is usually called water level section. For 3-axis milling, the water level section could have many loops, which causes complications in the set operation between sections. X and Y sections are single half loops and the Z value is unique for every points. Thus, it is important to have the set operation designed in more simple and effective manner.

A working system for using IPM in pre-form forging design is described in (Liu 1991, 1992). A drawing sheet with part sections is first created using the BACIS command language of CV/MEDUSA CAD system. Since there are many sections in a drawing sheet, each section wire-frame is assigned to a different layer according to its Y distance, and a certain number of sections can be looped through layers. Then each cutter section is moved to its cutter location for comparison with the part sections. The overlap between the cutter section and the part section will be removed from the part section. A real milling IPM can be obtained from the collection of the result sections. The display of sections is provided by line segments which may become confusing when there are too many lines. A realistic solution is to render the IPM as a 3D image. For this purpose IPM surface normal

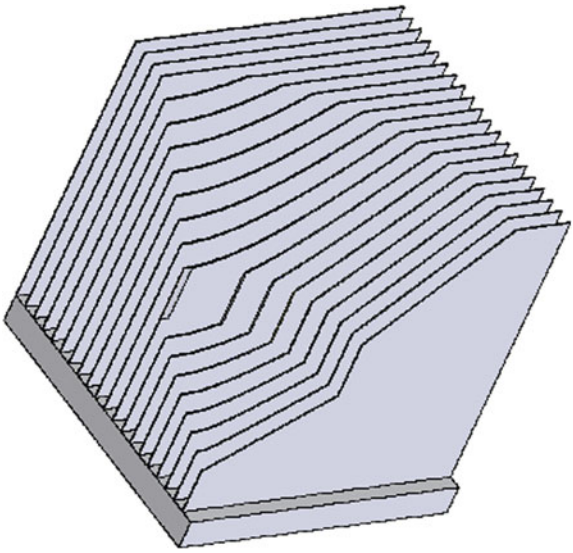


Fig. 3 Sectional representation

is needed for rendering computing. A sectional wire frame is divided along the X direction by the same step over of Y direction. A so-called regulated section is formed to facilitate the calculation of surface normal and interpolation of points between the sections. A given node in one section is linked to a node in the next section. A node's normal can be calculated from the four neighboring nodes. Figure 4 shows the regulated section representation.

The regulated section can also be used to accelerate set operation between cutter section and part section. Calculation of intersections and trimming between two sections are time consuming and the re-ordering of the line segments requires more computing time. This can be improved with the regulated sections, where the line segments are indexed by both cutter section and part section. Only the line segments with the same index are compared and trimmed. If all the line segments fall on the regulated nodes, there is no need to trim two line segments. The set operation can be simplified to the comparison of two Z values, which is very fast and stable. Hence, the Z map representation of IPM emerges (Jerard et al. 1989).

3.3 Z Map

If all the section line segments fall on the nodes, the object surface can be represented by the Z values of the nodes. A map of Z values represents the object geometry. In computer language terms, the Z map can be expressed as a two-dimension array Z[i, j], where i is the X index and j is the Y index. The XY position of the Z map can be calculated by i or j times grid size.

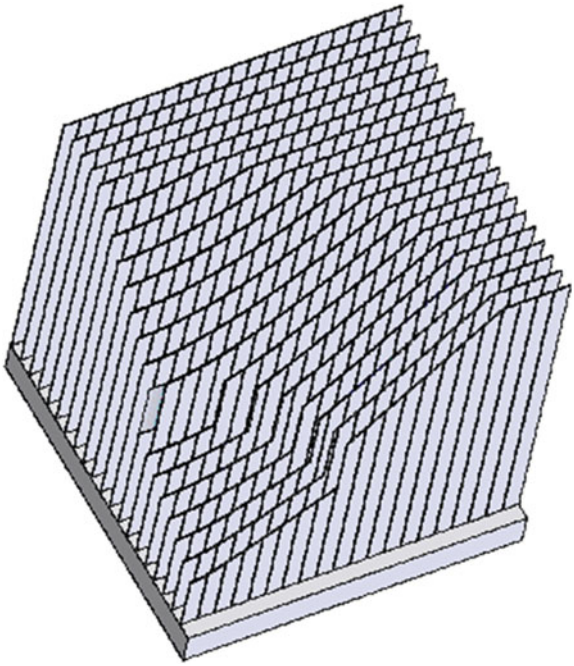


Fig. 4 Regulated section representation

The best analogy for a Z Map is a needle bed, where needles are uniformly distributed over the XY plane (Fig. 5). The tip of every needle touches the surface of objects. A milling simulation can be seen as the tool cutting through the needle bed. These needles can be described in mathematical terms as Z-axis aligned vectors standing on grid points on the XY plane. A Z Map representation can be used effectively to visualize the object surface defined on the XY plane. Since 3-axis milling parts are composed of surfaces visible from the Z direction, they can

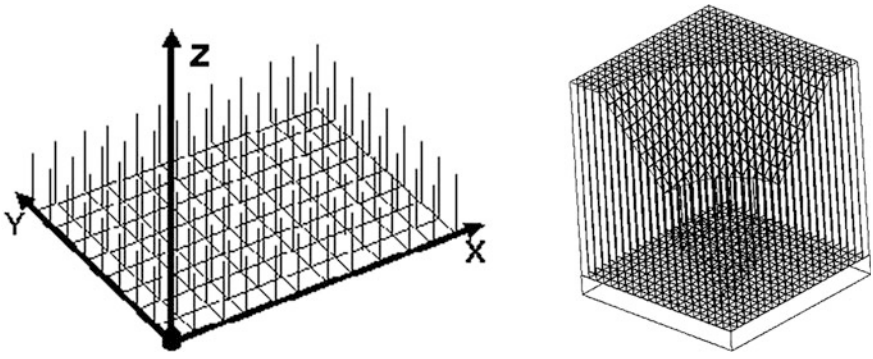


Fig. 5 Needle bed sample of classic Z Map model

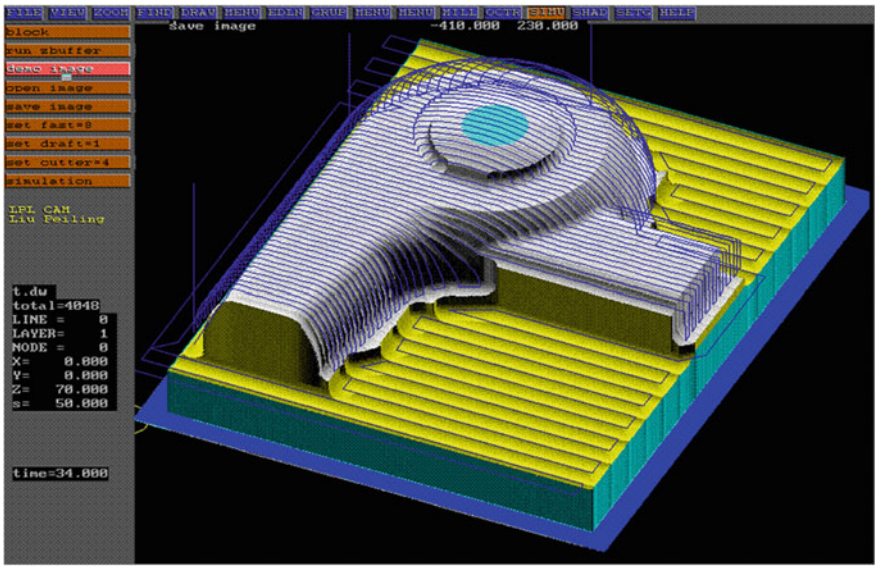


Fig. 6 Z Map IPM-based milling simulation

be expressed effectively by the Z Map representation. With a Z Map representation, the machining process can be simulated by cutting the Z Map vectors with the cutter. Figure 6 shows an example of 3-axis milling simulation. The vectors in a Z Map have direction and magnitude (length). The end points of all Z Map vectors form an approximate surface of the objects. Apparently, Z Map is a discretized representation of the object surface defined on grid points. For a non-grid point, it is possible to obtain its Z value of the point by simple interpolation based on the given Z Map. Various techniques (e.g., triangulation) are developed to render the Z Map model.

Obviously, the resolution of the Z Map (grid points) determines the precision of the Z map model. Z Map defined on finer resolution requires more intensive computing as well larger memory. A trade-off between the precision, computing load, and memory becomes a critical issue in Z Map. One of the simple strategies is to replace the common floating array by an integer array for Z Map which can reduce half of the Z Map size. Integer-based Z Map has an advantage to speed up the Boolean operation because the integer comparison is much faster than the floating one. The memory requirement of the Z Map is also halved by compressing the Z Map file. The Z Map model is widely accepted by commercial CAM software (Stifter 1995) due to its simplicity. Z Map model, however, has a problem in representing vertical wall due its slope requirement (Fig. 7). Z Map is perfect for forging die design as there are always draft angles in forging parts, but it may cause serious problems for milling parts since their profiling can often create vertical walls (Maenga et al. 2003).

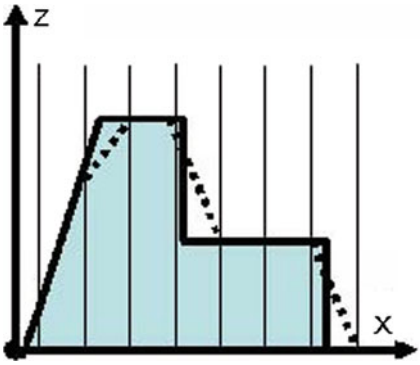


Fig. 7 Classic Z Map model with vertical walls

3.4 Extended Z Map

Since the precision of the Z map is determined mainly by XY grid resolution, increasing the resolution along the vertical wall adaptively is a key issue. Fortunately, it can be done by leveraging one important feature of 3-axis milling. Viewing from the top, the vertical walls usually cover only a small portion of the Z direction projection, so it is possible to use finer resolution along the vertical walls while maintaining a rough resolution in the planar area. This was the initial idea for an extended Z Map. Figure 8 illustrates the plan view of the Z Map grid.

The size of the grid can be reduced through using sub-cells. But the precision of the XY dimension is still limited by the size of sub-cells. For a sub-cell of 0.1 mm, the best precision is 0.1 mm in XY plane. There is a need to represent XY dimensions precisely. Instead of using vectors in the sub-cells, we use sticks in the sub-cells that have volumes and surface geometry. A B-rep surface model can be represented precisely using a Map of B-rep sticks (Fig. 9).

Milling simulation with stick method involves Boolean operation between cutter and stick. Figure 10 shows different stick shapes after cutting. The experiments with B-rep stick model are very slow as a huge B-rep model will be created using this method. To simplify stick and Boolean operation, a polygon

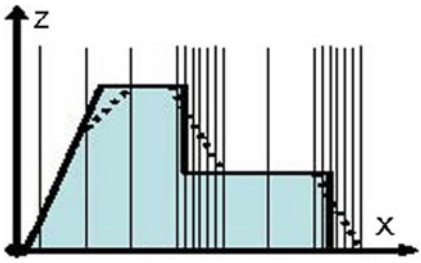


Fig. 8 Extended Z Map with sub-cells along vertical walls

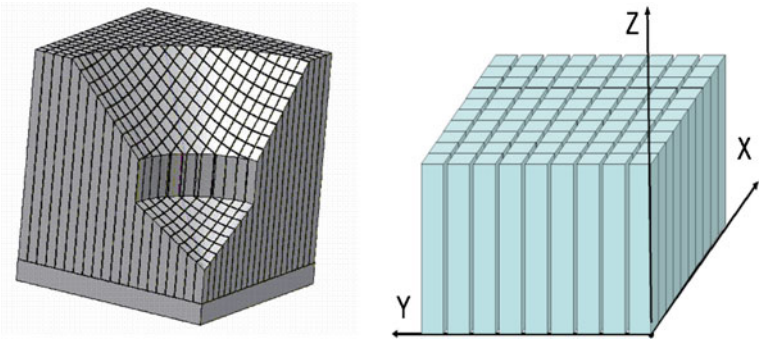


Fig. 9 Stick method

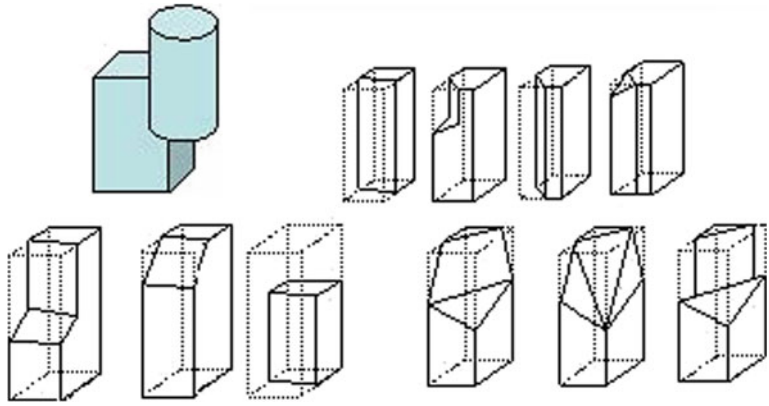


Fig. 10 Different shapes of stick elements

representation is used instead of real surface in a stick cell. The data structure of polygons is much simple than that of a B-rep which needs a group of complicated pointers to maintain a double wing data structure. Figure 11 shows a shop floor example of extended Z Map IPM-based NC simulation and verification (Liu et al. 2002).

It is a common practice to machine on six sides by flipping workpiece on 3-axis milling machine. However, the extended Z Map has limitation on this situation, as the stick representation of the workpiece can only simulate 3-axis machining in a fixed orientation.

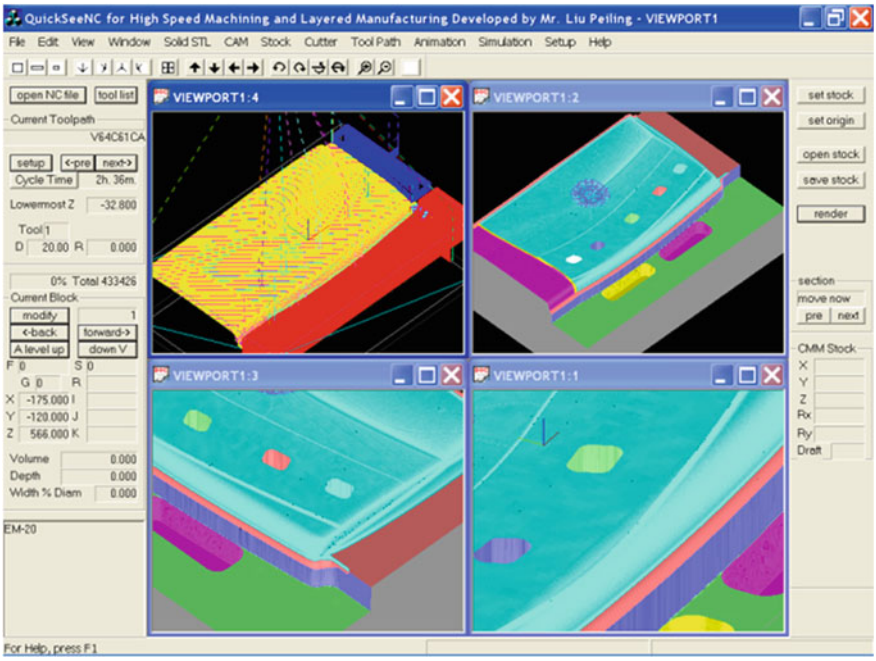


Fig. 11 Extended Z Map for IPM-based NC simulation

4 Voxel-Based In-process Model for Virtual CNC Training System

The term voxel represents a volume element in 3D space to decompose geometrical models by regular and simple substructure. It is an extension of 2D pixel representation denoting a picture element in raster graphics. Voxelization is the process of converting a 3D object into a voxel model. Volume graphics, voxelization, and volume rendering have attracted considerable research works in recent years. However, all of this work is directed at the display of volume data, mainly for medical applications. In this chapter, a simplified voxel-based IPM is proposed to simulate the multi-machining processes. Figures 12 and 13 depict the voxel model and its implementation in machining simulation, respectively. The preliminary results have shown that the voxel-based volume modeling is a very promising approach to the unified IPM for multiple machining and layered manufacturing simulations.

Voxel-based representation has an advantage in dealing with heterogeneous material. The memory requirement for traditional voxel models, however, is enormous. Storing the voxel array in compressed form and developing algorithms to operate directly on the compressed data is an interesting area of research. A voxel-based system should be able to update the display in real-time. To achieve this, parallel algorithms and hardware support for volume rendering are the focus of the current

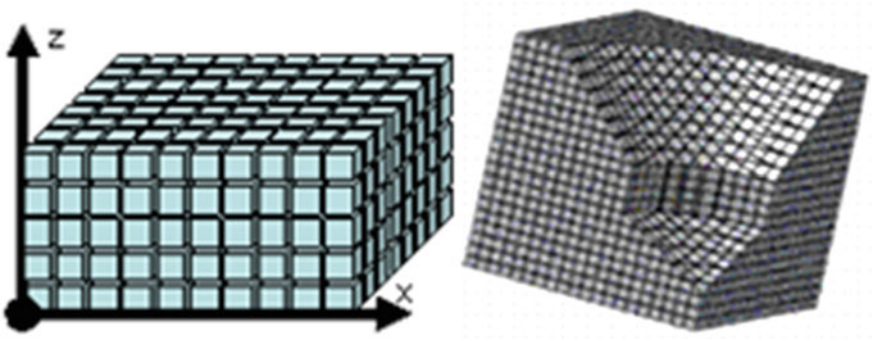


Fig. 12 Voxel model

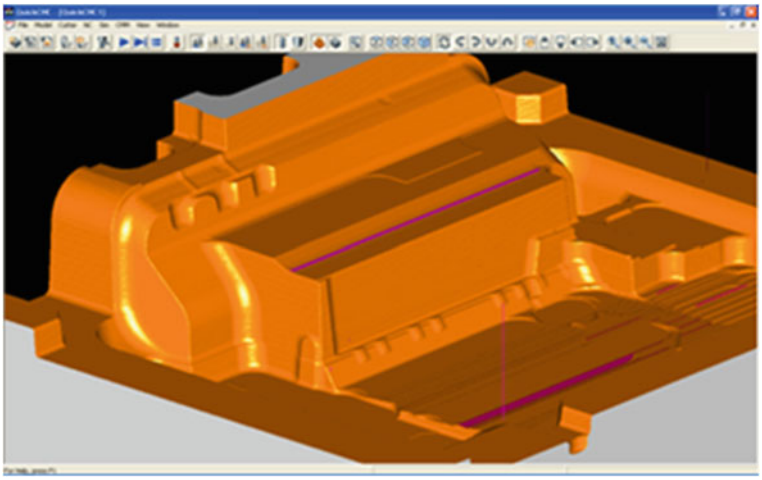


Fig. 13 Voxel model implementation in machining simulation

research. We only render boundary voxel to effectively visualize the model in real time avoiding latency problem of expensive ray-casting with huge internal voxels. This can be easily achieved by rendering of a points cloud with the boundary voxels.

5 Industry Applications of Virtual CNC Training System

The proposed new IPM has been used in the virtual CNC training system developed in Singapore Institute of Manufacturing Technology (SIMTech) for training of CAM programmer and CNC machinists. The machine frame, dial indicator, and control panel of the system are shown in Fig. 14. The trainees can use the system to simulate the milling process and save the "machined" model for other downstream machining processes. Figures 15 and 16 demonstrate the simulation of the

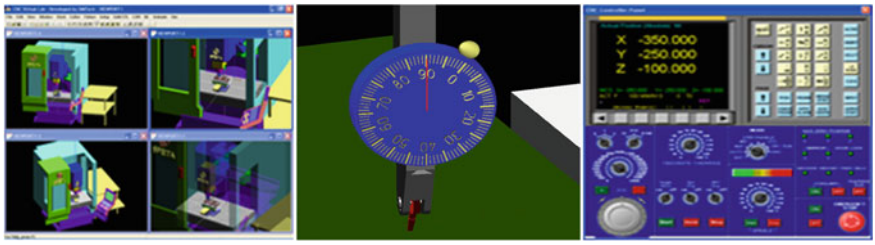


Fig. 14 Machine frame, dial indicator, and control panel

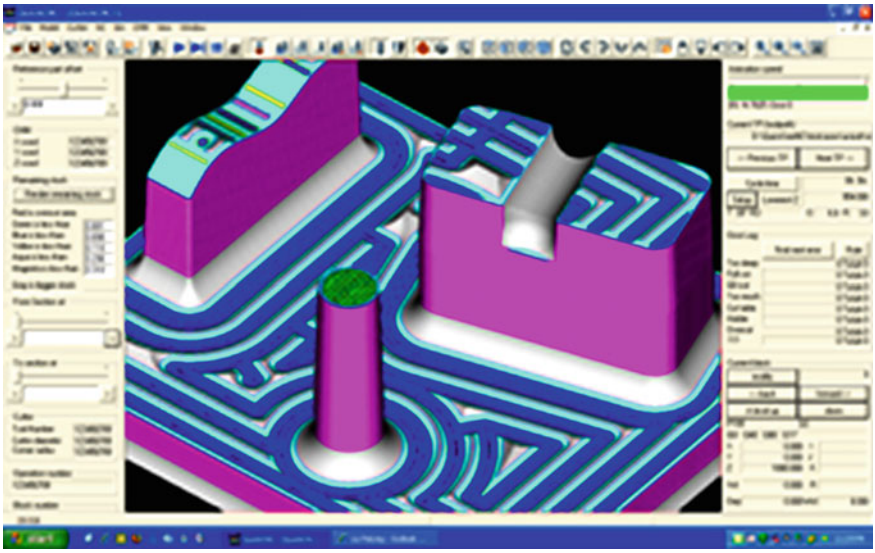


Fig. 15 Virtual CNC simulates the remaining of stock

remaining stock and the scallop height respectively. In addition, they can control the simulation speed to see the details at any angles on the current situation of the machining, which is difficult if not possible in the real machine-based training.

A set of different machining samples has been provided to demonstrate how the generated tool-path works with cutter under various cutting parameters with the aid of the virtual controller. Trainees can learn different setups in a short time using virtual simulator on PC, which significantly shorten the learning curve compared to the traditional training in a workshop.

While it is difficult to show the effect of a wrong setup or NC code on shop floor, the virtual simulator has graphics and sound developed to synthesize various effects. In particular, the virtual CNC training system can simulate an accident using a graphic and sound effect when a trainee breaks a leg of a workpiece during the virtual machining. Among several benefits safety and material cost reduction are the direct and major gains for the virtual simulation.

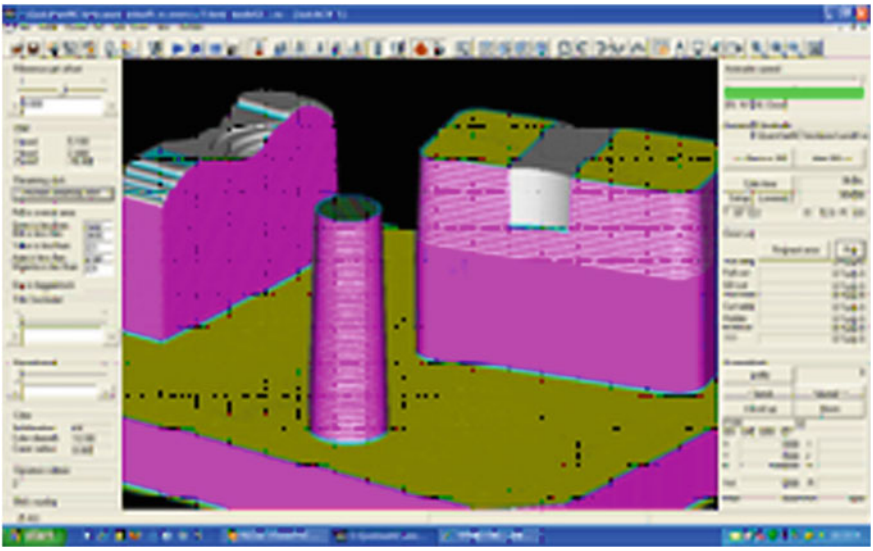


Fig. 16 Virtual CNC simulates the scallop height

The Precision Engineering Worker Skill Qualification (PE WSQ) Specialist Diploma is a joint initiative by SIMTech and the Singapore Workforce Development Agency (WDA) to provide hands-on training to equip future PE professionals in cutting edge precision machining processing technologies. This program is conducted through a series of lectures, laboratory demonstrations, and project attachments in selected industrial applications. As most of the training organizations have limited numbers of CNC machine tools and CNC trainer available, they can install the virtual CNC training system on their PCs to conduct hands-on training. In this program, the virtual training laboratory is designed for 40 students to learn CNC. High speed machining course is conducted for the WSQ trainees to learn machining using the system. Twelve sets of HSM examples are created and allowing trainees to learn different machining techniques and strategies, one of them is shown in Fig. 17. By using Virtual CNC Training, it can effectively reduce CNC learning curve from typical weeks long to just one night. Trainees can do self-learning using the same software in their own PCs.

With the financial support from Local Enterprise and Association Development Programme (LEAD), SPETA has deployed the virtual CNC training system in their classroom. As one of the several critical areas, they identified to enhance the capabilities of the PE companies; training of CNC machinists has paid great attention by them. In partner with SIMTech, SPETA uses the system to train CNC machinists—somewhat like the flight simulators to train pilots, which will significantly trim the training hours on the actual machine (Fig. 18).

The system is also used in Institute for Technical Education (ITE) for training computer numeric control machinists. Significantly reducing the hours and

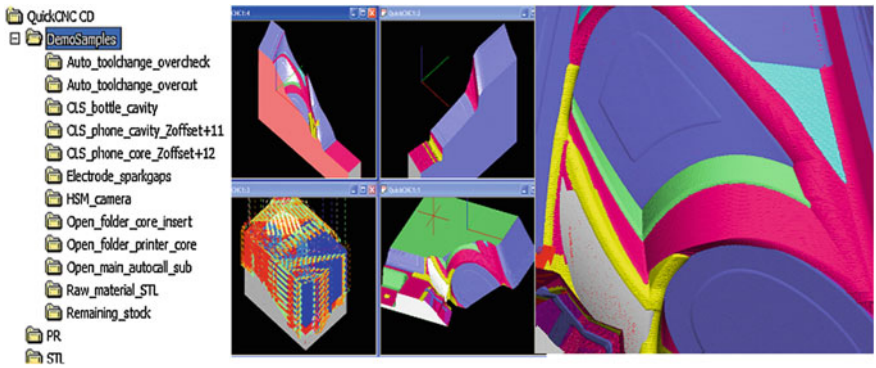


Fig. 17 Virtual CNC machining example for WSQ course

Fig. 18 Virtual CNC Training has been implemented in SPETA

Fig. 19 Virtual CNC in MTA and Science Festival 2011

machine resources required, the virtual CNC training system enables trainees to practice more with various machining requirements within the same allocated training time. With this additional preparation, trainees would have a shorter learning cycle when they start working with the companies. Virtual CNC Training has been featured in local TV, Radio, and all newspapers. Singapore Science Festival 2011 and Metal Asia (MTA) both hold daily shows for Virtual CNC Training (Fig. 19).

6 Conclusion and Discussion

After many years of research and development, SIMTech has designed and developed the world's first virtual CNC training system using a three-dimensional precise unified geometrical model. The system allows trainees to simulate the milling process and save the "machined" model for other downstream machining processes. It also allows for different situations to be tested during training, which would be costly if done on the machines. It is more engaging to operate on virtual machines than conventional classroom-based lecturing.

The lack of haptic or tactile feelings for dynamic simulation is one of the limitations with the virtual CNC training system. For example, during the real machining, cutter always vibrates under different cutting conditions. Long tool may chatter under too deep cuts, which will generate marks on workpiece surface. Cutter wears off during long time of cutting.

Future developments include extending this VR concept from CNC to other machining operations. We are interested in development of more realistic machining training that can simulate tool chatter, cutter deflection, and tool wear in near future.

References

Avgoustinov N (2000) VRML as means of expressive 4D illustration in CAM education. Future Gen Comp Sys 17:39–48

Chandru V, Manohar S, Prakash CE (1995) Voxel-based modeling for layered manufacturing. IEEE Comput. Graphics App 15(6)

Donggo J, Kwangsoo K, Jungmin J (2000) Voxel-based virtual multi-axis machining. Int J Adv Manuf Technol 16(10):709–713

IMTI, Inc. (2002) Unit Process Modeling for Affordable Manufacturing, White Paper, www.imti21.org

Jerard RB, Hussaini SZ, Drysdale RL, Schaudt B (1989) Approximate methods for simulation and verification of numerically controlled machining programs. The Visual Computer 5(4):329–348

Liu PL (1991) A new concept integrated CAD/CAM system for complicated die and mold. Advances in Computer Science Application to Machinery 8:90–95

Liu PL (1992) 3D complicated parts design based on thes automatic shape generation. Chinese Journal of Mechanical Engineering (English Edition) 5(2):88–92

Liu PL et al (2002) An object representation method, WO04032001A1

Maenga SR, Baekc N, Shinb SY, Choid BK (2003) A Z-map update method for linearly moving tools. Comput Aided Des 35:995–1009

National Academy of Engineering, http://www.engineeringchallenges.org

Ong SK, Mannon MA (2004) Virtual reality simulations and animations in a web-based interactive manufacturing engineering module. Computers and Education archive Volume 43(4):361–382

Park SC, Mukundan G, Gu S et al (2003) In-process model generation for the process planning of a prismatic part, J Adv Manuf Sys 2(2):147–162

Qiu ZM, Chen YP, Zhou ZD, Ong SK, Nee AYC (2001) Multi-User NC Machining Simulation over the WWW. Int J Adv Manuf Technol 18:1–6

Stifter S (1995) Simulation of NC machining based on the dexel model: a critical analysis. Int J Adv Manuf Technol 10(3):149–157

Updated GameTools: Libraries for Easier Advanced Graphics in Serious Gaming

Rubén Jesús García, Jesús Gumbau, László Szirmay-Kalos and Mateu Sbert

Abstract Advanced graphics effects can be difficult to create and implement by nonexperts. We present here multiplatform, multigame engine libraries, designed to simplify the use of efficient state-of-the-art computer graphics algorithms in the fields of geometry and illumination. A summary of our experiences updating and porting the libraries is included, with recommendations possibly useful to other programmers. The main features of the updated libraries are described. Two serious games which use the geometry libraries are presented here as use cases, both dealing with teaching history to middle school students. Finally, we show an implementation of a board game played in Europe since roman times (of interest to museums specialized in medieval or roman times) which uses the illumination libraries.

Keywords Game engine · Serious game

R. J. García (✉)
Institute of Informatics and Applications, University of Girona, Girona, Spain
e-mail: rgarcia@ima.udg.edu

J. Gumbau
Universitat Jaume I, Castellón, Spain
e-mail: jgumbau@lsi.uji.es

L. Szirmay-Kalos
Budapest University of Technology and Economics, Budapest, Hungary
e-mail: szirmay@iit.bme.hu

M. Sbert
University of Girona, Girona, Spain
e-mail: mateu@ima.udg.edu

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_12,

1 Introduction

A serious game is a game designed for a primary purpose other than pure entertainment. As a consequence, the development team is often more centered on realistic modeling of the aspects of the virtual world which are required for the teaching of techniques which the user will later use in his professional life.

It is common that designers and programmers are experts in the application domain, rather than in the field of computer graphics (especially when comparing to firms creating entertainment games).

The creation of new graphic effects is becoming increasingly complex, due to two main reasons:

- New effects often require the use of advanced capabilities of Graphical Processing Units (GPUs), which use architectures quite different from that of CPUs, and require the knowledge of specialized languages (Cg, HLSL, GLSL, CUDA) and techniques (parallel architectures with distributed memory, specific geometry, vertex and pixel shaders, and different types of memory with tricky access modes).
- These effects are becoming increasingly supported by complex physics simulations, which require knowledge of methods to solve differential equations efficiently.

Therefore, the existence of a general library of graphic effects and techniques, which could be used in different projects by nonspecialists in the field of computer graphics, can be a great asset for researchers and programmers in the serious gaming community.

The next section describes previous work on which our library is based, while Sect. 3 describes the new libraries' porting procedure and implementation details. Section 4 contains three use cases: the Jaume I, Legends of Girona, and Nine Men's Morris games. Finally, Sect. 5 concludes the chapter.

2 GameTools

GameTools (2008b) was a European Union project which brought together six universities and four industrial partners during 2004–2008, and whose aim was to create a set of *graphic libraries* encompassing different aspects of computer graphics: geometry simplification, global illumination, and visibility culling.

The libraries were meant to aid game companies who did not have a large enough budget to maintain a specialized team in advanced computer graphics techniques. Companies were supposed to focus on the game story, ensuring playability, and user enjoyment, while the advanced graphic effects were provided by the GameTools libraries. GameTools libraries were provided for the windows platform, using visual studio 2005 and DirectX 8, with support for the Ogre3D (OGRE 2012) game engine, version 1.0.8, and the Shark3D (2012) engine.

2.1 Geometry Libraries

To create a virtual world, one of the tasks is modeling scenarios and characters using geometric primitives, usually triangles. These (usually complex) models require substantial processing power to render. In order to decrease computation time without sacrificing quality, the geometry of far-away objects can be transformed (simplified) so that the cost of rendering is significantly reduced. Since the simplified objects occupy a small area in the screen, imperfections are difficult to distinguish to the human eye. If the level of detail is changed continuously as the object moves, the change will not be noticed by users.

The Geometry libraries contain algorithms dealing with geometry processing in the Context of Continuous Level of Detail (CLOD). In particular, mesh simplification (González et al. 2007b), spherical light fields (Domingo et al. 2007), tree leaves (Rebollo et al. 2007a, b), and mesh management (Gumbau et al. 2007) are included. Error metrics taking textures into account are also used (González et al. 2007a).

An example of the software developed to manage simplification methods can be seen in Fig. 1.

2.2 Illumination Libraries

The illumination libraries contain an assortment of efficient algorithms for realistic rendering, based on the concepts of approximate raytracing (Szirmay-Kalos et al. 2005a, b; Umenhoffer et al. 2007; Lazányi and Szirmay–Kalos 2005) (Fig. 2), indirect illumination gathering (Méndez et al. 2005; Szécsi et al. 2006; Lazányi and Szirmay-Kalos 2006; Szirmay-Kalos and Lazányi 2006; Umenhoffer and Szirmay-Kalos 2007) (Fig. 3), scattering media simulation (Umenhoffer et al.

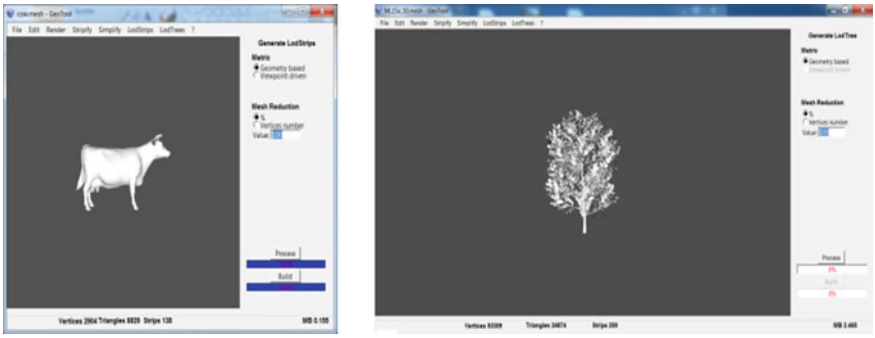


Fig. 1 Continuous level of detail: *triangle* strips for general models; *tailored* algorithms for vegetation

Fig. 2 Approximate raytracing: realistic transparency and caustics

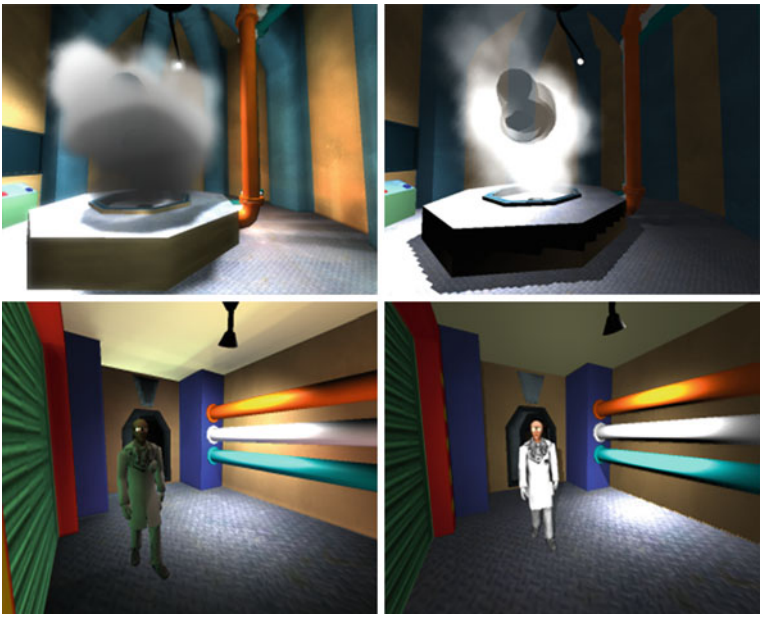


Fig. 3 Indirect illumination gathering (*left*) versus local illumination model (*right*)

2006; Umenhoffer and Szirmay-Kalos 2005, 2006) (Fig. 4), and postprocessing effects such as depth of field, tone mapping (Toth and Szirmay-Kalos 2007), glow (Fig. 5), and image distortion caused by heat shimmering (Fig. 4 left, and Fig. 6). The effects can be used together with no restrictions.

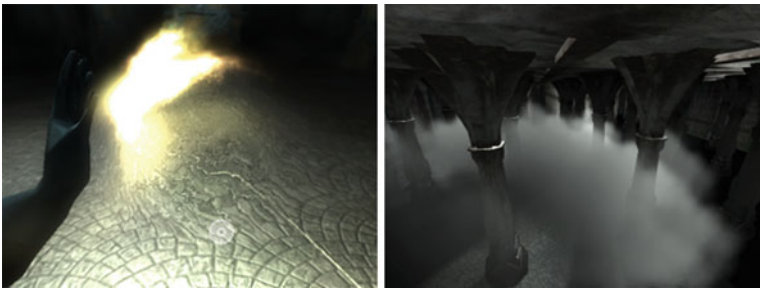


Fig. 4 Scattering media simulation: fire and smoke

Fig. 5 Postprocessing examples: depth of field, tone mapping

Fig. 6 Postprocessing examples: image distortion due to heat

Fig. 7 Potentially visible set (PVS) and coherent hierarchical culling (CHC)

2.3 Visibility Libraries

Given a scene that we wish to render, the visibility problem consists in calculating efficiently which geometry is (at least partially) visible from a viewpoint (Hadwiger and Varga 1999). If we have some information about the type of scene being rendered, we can sometimes apply specific algorithms to discard invisible geometry.

The GameTools visibility libraries contain state-of-the-art visibility culling algorithms, examples of which can be seen in Fig. 7. Figure 7 (left) shows an interactive walkthrough of the city of Vienna which uses a potentially visible set visibility culling algorithm. Figure 7 (right) shows the use of coherent hierarchical culling (Bittner et al. 2004) to reduce the number of rendered objects rendered from the initial 2000 to only 425.

3 Updated GameTools

A new project was started in 2009 to update the geometry and illumination routines of GameTools to support other operating systems and gaming platforms. Since the Ogre3D engine is multiplatform, running on windows, linux, and OSX, it was considered a worthy endeavor to rewrite the GameTools routines portably, using the Ogre3D engine as a test bed to check for regressions.

However, the rapid pace of development in free software projects, when compared to proprietary software (and the lesser emphasis on backwards compatibility) have forced us to follow the Ogre3D development closely and continue updating our libraries at a fast pace. Currently, our version uses Ogre3D 1.7.3 and Unity3D 3.4. A simplified structure diagram of the routines can be seen in Fig. 8.

The illumination routines are divided into render techniques (materials) and render runs (auxiliary techniques used by the render techniques). In Ogre3D and Unity3D, materials are built using render techniques. Unity3D scripts control some

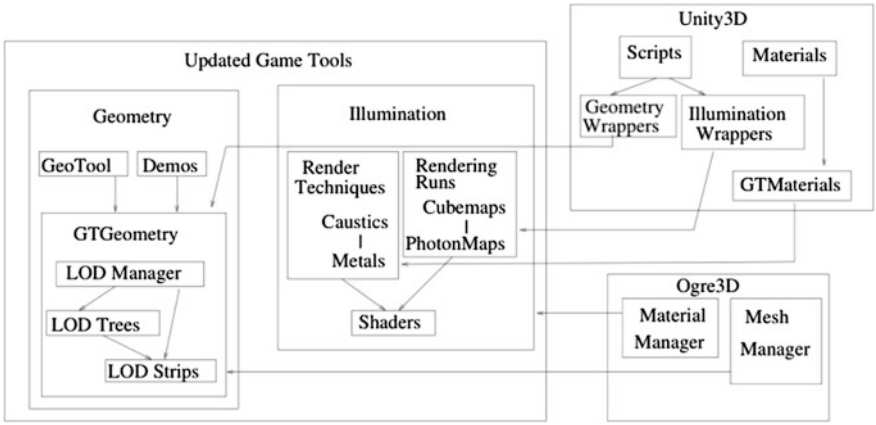


Fig. 8 Structure diagram of the updated GameTools libraries, and their integration in the Unity3D and Ogre3D engine

of the functionality. The geometry routines control the LOD of the models, and are called from the material manager in Ogre3D and from Unity3D scripts.

3.1 Porting Procedure and Lessons Learnt

The original GameTools libraries follow correct design, programming and testing procedures, and work quite well in their supported platform. However, while updating the code, we have found undocumented assumptions which introduce hidden bugs and make porting more difficult. This section describes some of these bugs and describes procedures which would have found most of these bugs in the early phases of development, when they could be fixed more easily.

3.1.1 Persistent Data Structures

The GameTools libraries create persistent data for different purposes. For example, the geometry library creates and stores triangle strips and additional data to manage the CLOD of both triangle strips and procedural trees. The illumination library creates and stores illumination samples and data for the precomputed path maps.

These files contain different types of integers, floating point, and boolean data, and since the only supported platform was Visual Studio 2005 on Microsoft Windows, the files were created by storing the binary data directly onto disk, and read by transferring the data from disk into memory. No effort was spent documenting endianness, size of data types, alignment padding or otherwise. The move

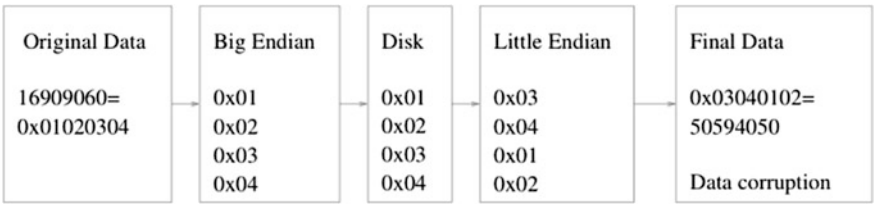


Fig. 9 Example of file transfer between machines with different endianness creating corrupted data

to Windows on 64 bit architectures was not foreseen. This means that the use of a different compiler (even on 32 bit Windows) would introduce bugs in file access, and complicates the porting to Mac OSX (where the size of booleans is different), 64 bit Windows and linux (where the size of long is different) and tablets (where endianness is different). See Fig. 9 for an example.

This difficulty in porting could have been prevented if data types had been standardized in the software design phase, and if proper methods to separate memory data structures from disk data structures had been added. We recommend adding a variety of hardware and software platforms to the testing procedures, including big and little endian architectures, different word sizes if possible, and different compilers, even if only one platform will be eventually supported. This ensures that nonportable code can be detected early, isolated if the effort to make it portable is deemed too high, and that file formats are properly documented.

3.1.2 Memory Access Problems

Memory access problems are the most common bugs in C++ development, and many applications have been created trying to reduce this problem, even going to the trouble of emulating the whole processor and memory access (Valgrind 2012; Seward and Nethercote 2005). Despite that, mistakes in memory access remain in most programs, and are a major fraction of crashing bugs and hacking/cracking attempts.

Our experience porting and updating the GameTools routines showed that these programming errors can remain hidden and be difficult to fix because of two reasons. First, the memory management code on Microsoft Windows does not reuse free memory very aggressively, so freed memory remains ‘valid’ for potentially a large amount of time. Second, the Visual Studio 2005 implementation of the standard template library collections in many cases does not take advantage of the fact that references to objects in collections are not valid after a change (this would allow the use of complex data structures to decrease computation time). Software testing does not find any of these errors, since to the operating system, the memory is actually valid. Again, testing on different architectures, compilers and operating systems permits early detection and fixing of these errors at a small cost.

3.1.3 External Libraries

Complex software is not created from scratch. In most cases, many already existing libraries are reused, and care must be taken to write efficient code to interface with them. This is especially problematic when libraries expect data in different formats, since in many cases more time is spent translating formats than doing actual work.

However, the main problem we have encountered is that many libraries which appeared to be thriving during the time frame of the original GameTools project have been abandoned, forcing us to either use old, unmaintained versions with unpatched security problems, or "inherit" these libraries and update them ourselves. The problem with this last approach, of course, is that the managing of large libraries is a big project in itself, and managing more than one can be overwhelming for small groups.

This problem can be mitigated by identifying the supporting libraries, finding other libraries solving similar problems, and designing an abstract interface which can deal with different back-ends (the model view controller architecture is one possible architecture). Although this is more costly than choosing only one library, the long-term benefits usually are worth the cost. Many large libraries, therefore, follow this approach, which we recommend. An example is OpenCV (2012), which is one dependency of our augmented reality demo. See Fig. 10 for an overview.

Interestingly, many well-maintained libraries presented the opposite problem: development was so fast that retaining compatibility with updated versions of the

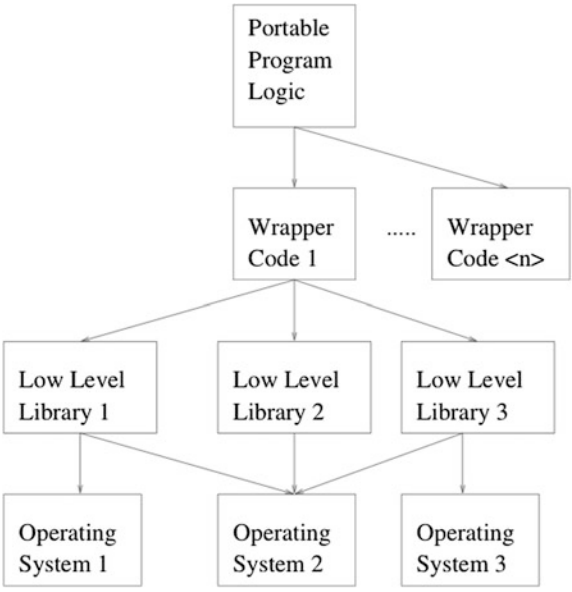


Fig. 10 Example of runtime or compile-time selection of libraries in portable code

libraries was much more costly than originally expected. In particular, the Ogre3D game engine released two major versions (and eight minor versions) during the time frame of our new project, which required updating of the interfaces between the GameTools libraries and Ogre3D.

This last problem is quite important when fusing free software libraries (which seem to have a quicker pace of development), since the support for older versions from the development team is small, and since having access to the source code encourages the changing of the APIs and the removal of older, less efficient APIs, which breaks binary compatibility.

Additionally, care should be taken when choosing external libraries to preserve the portability of the whole system. Not only should libraries advertise support for the operating systems targeted by the software being developed, but testing procedures should be added to test this support, since in many cases, support can be enough to create simple applications, but advanced features are not supported. For example, although Unity3D supports both Windows/DirectX and OSX/OpenGL platforms, the OpenGL support only supports second generation hardware, and does not permit the use of complex, third or fourth generation shaders, even though the underlying operating system and card supports it. As a consequence, our advanced illumination routines cannot run on OSX Unity3D.

Another caveat worth mentioning is the licensing of these libraries. While libraries are chosen taking their licensing into account so that they can be bundled in final products, it is sometimes the case that the licensing of the final product changes in order to respond to marketing or other pressures. In our case, the original GameTools libraries were created under european union funding, and original plans included both open source licenses (for initial libraries) and commercial licenses (for optimized code) (GameTools 2008a). The follow-up project was started under a Catalan grant, and encouraged proprietary licenses.

These changes in licensing require a reassessment of all the supporting libraries, and may require new libraries to be swapped in. Though the approach presented above will help the transition, we cannot encourage enough contacting the copyright holders of all used libraries beforehand and requesting their opinion on different types of licenses (including obtaining customized licenses).

3.1.4 Portability in GPU Code

We have used Nvidia's Cg language to create portable algorithms running on GPUs. Cg can be used from both OpenGL and DirectX, and runs on windows, linux and OSX; the GPUs supported include nVidia, 3Dlabs, ATI and Matrox. Additionally, Cg and HLSL are quite similar languages, allowing us to reuse the original HLSL code created for the original GameTools libraries.

Nevertheless, we have encountered some small, unexpected portability problems, which we detail here. Although Cg allows us to write portable programs, in many cases some parts of the graphic pipeline environment implicitly affects the result. The default values of the projection and modeling matrices, and other

rendering state, are different in DirectX and OpenGL. As a result, even simple code might need to be either rewritten to take these values into account, or the expected value of all the rendering states affecting the GPU code must be documented. In particular, since the handedness of OpenGL and DirectX is different, checking the z-buffer might require changing the comparison operators. This is complicated by the fact that Ogre3D uses an OpenGL-like environment when rendering under DirectX. Additionally, the use of external rendering engines and libraries restricts the constructs which can be used in the Cg code, since rendering engines may have old versions of the Cg compiler, or use HLSL compilers with additional restrictions on the code they accept.

We again recommend testing all available architectures from early stages of development, when all the implementation details and decisions are still fresh in the developers' minds.

3.2 Supported Game Engines and Libraries

The updated GameTools libraries are portable, running on Linux, Windows, and Mac OSX. Supported game engines include Ogre3D and Unity3D (Unity 2012), but users may use them in conjunction with other engines or libraries. Example C-language wrappers and different demos showing the use of the API are provided. An example of an augmented reality demo using the ArUco (2012) and OpenCV libraries is shown in Fig. 11.

3.3 Geometry Library: GTGeometry

The GTGeometry library contains different modules dealing with CLOD algorithms:

Fig. 11 Example augmented reality application using our libraries, running on GNU/Linux and OSX, at different LODs

- Stripification: takes a list of triangles as input and produces triangle strips suitable for efficient rendering.
- Simplification: generate geometry or viewpoint-driven simplified meshes.
- Simplification of trees: specific algorithms for vegetation.
- Managing of LODs: Automatic calculation of optimal LODs.

The library is backwards compatible with the original GameTools geometry libraries, and has been ported to the Unity3D engine. The modules are described in more detail in the following sections.

3.3.1 Stripification Module: LODStrips

The stripification module contains code to create efficient triangle strips from general models. The first demo demonstrate the LODStrips multiresolution run-time library (Geometry: LODStrips Library). The application (Fig. 12) shows a group of models which are able to change their level of detail depending on the distance of the group of objects to the camera. The information panel on the bottom-left corner of the screen shows the current LOD factor, frames per second and the amount of geometry sent to the renderer. It can be seen how the frame rate increases as the LOD decreases.

The level of detail can be calculated in two ways: automatic LOD (based on the distance of the group to the camera) and manual LOD (the user changes the level of detail independently from the distance), which changes the level of detail of the objects manually. This last mode is useful to see the meshes in detail even when their level of detail is set to the minimum.

The demo also shows how LOD operations can be minimized by grouping some instances of the same model to be managed by a single LODStrips multiresolution model. This is useful when some models need similar levels of detail and will improve the overall performance.

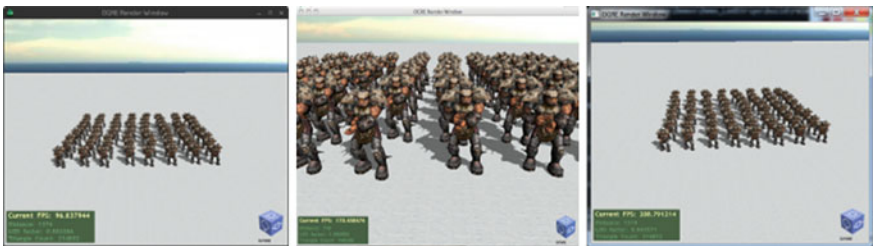


Fig. 12 LODStrips example, Ogre3D, Linux, Mac OSX, Windows

3.3.2 Continuous LOD Managing Module: LODManager

The LODManager demo features a massively populated scene composed of 1,200 models. Each one of them is attached to an independent multiresolution model instance that manages its level of detail. To manage the level of detail of such a vast scene, we introduce the use of the LODManager, which decides whether an object can change its level of detail freely or just has to borrow an already calculated LOD snapshot.

The demo allows the user to enable or disable the LODManager capabilities to show the difference in performance. The LODManager is able to keep the bottleneck of the application in the graphics engine and not in the LOD calculations. A screenshot can be seen in Fig. 13.

3.3.3 Continuous LOD for Vegetation: LODTrees

This demo presents Geometry: LODTreeLibrary, the LODTrees multiresolution runtime library. The demo is composed of some groups of multiresolution trees. The LOD of each group, composed by some trees of the same type, is managed by a single LODTree instance to optimize performance. The result is a forest of multiresolution trees which change its level of detail depending on the distance to the camera of each one of these groups.

The geometry of the trees in our demo is provided by the Xfrog software (2012), but the LODTrees module itself can use geometry from other sources. Figure 14 shows an image of the LODTrees demo.

3.3.4 Geometry Management Application: GeoTool

GeoTool is an application which eases the creation of the LOD models. It uses the FLTK (2012) toolkit to ensure portability among the three supported architectures. Its main features include simple operations such as triangle stripification, mesh

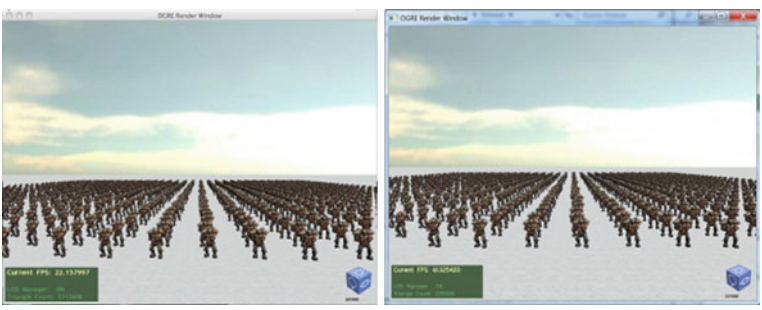


Fig. 13 LODManager example, Ogre3D, Mac OSX, Windows

Fig. 14 LODTrees example, Ogre3D, GNU/Linux and Mac OSX

simplification (geometry or viewpoint based), leaves simplification, and complex operations (e.g., LODStrip and LODTrees creation and visualization). A screenshot of the software can be seen in Fig. 15.

3.3.5 GTGeometry in Unity3D

Unity3D Pro allows native C code to be called from within its scripts. To use GTGeometry within Unity3D, C wrapper routines can be used. An example of a LODTree controlled within Unity3d can be seen in Fig. 16.

3.4 Illumination Library: GTIllumination

The GTIllumination library contains different global illumination effects, implemented efficiently using GPU programming. Effects include realistic metals and glass, caustics, particle systems, hierarchical systems, tone mapping, precomputed radiance maps, multiple reflections and refractions, and various screen space

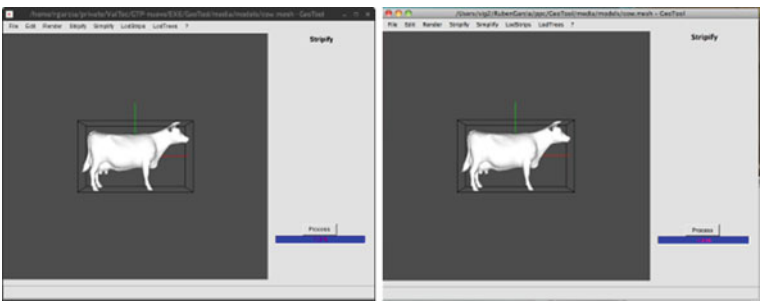


Fig. 15 GeoTool application, GNU/Linux and Mac OSX

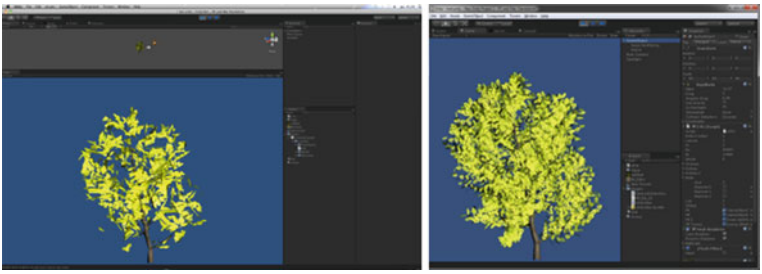


Fig. 16 Continuous LOD Tree in Unity3D, Mac OSX and Windows, at a LOD of 30 % (*left*) and 100 % (*right*)

ambient occlusion algorithms. Figures 17, 18, 19 and 20 show screenshots of the different effects.

The software package contains GPU shaders used to implement the different materials and effects. Interfacing packages have been created for Ogre3D and Unity3D in order to make the effects easy to use in these engines. The work flow is similar to using predefined materials and effects. Ogre3D's materials and techniques can be defined in material scripts with the use of keywords, and a Unity3D package is provided to add the new functionality, so that drag-and-drop can be used to indicate techniques and materials.

Scripts and prefabs are provided to link the different techniques to the relevant Unity3D objects; the only caveat is that since the (color and distance) cubemap textures required to calculate accurate reflections are included in the material, each

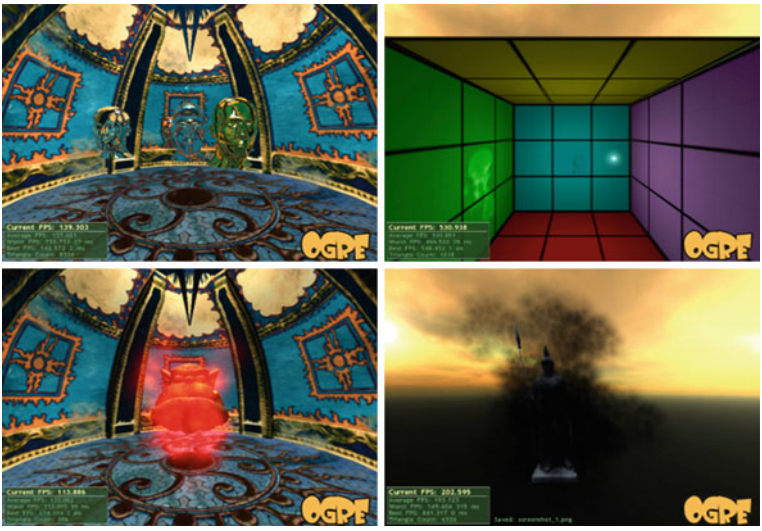


Fig. 17 Realistic metals and glass, caustics, particle systems, hierarchical systems

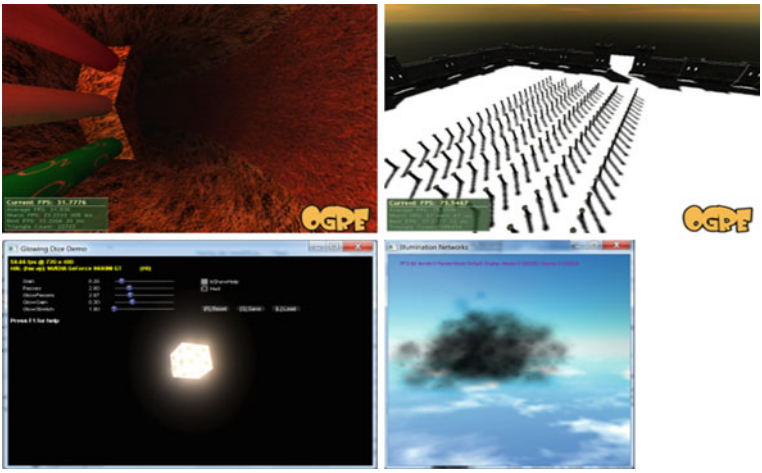


Fig. 18 Precomputed radiance maps, soft shadows, glowing effect, and illumination networks

Fig. 19 Tone mapping and eye adaptation

object should have its own instance of the material (the material can be shared among nearby objects but the quality of the picture will be affected, since objects sharing materials will not be reflected on each other). Figure 21 shows an example of the Unity3D package in action. The material for the skull object is a realistic gold, and its parameters and textures are updated in real time by a child CCM prefab containing the scripts and cameras required to update the color and distance cubemaps and other parameters. Additional scripts (not shown) are attached to the camera to ensure correct image generation.

Figure 22 shows examples of metal materials on the windows platform. The effects have been confirmed to work in Linux and Mac OSX using the windows emulator Wine (2012) (see Figs. 23 and 24), so we are confident that our routines will be available in Unity3D on the Mac platform as soon as Unity3D updates their OpenGL support, and on the linux platform when Unity3D adds support for the platform in the near future.

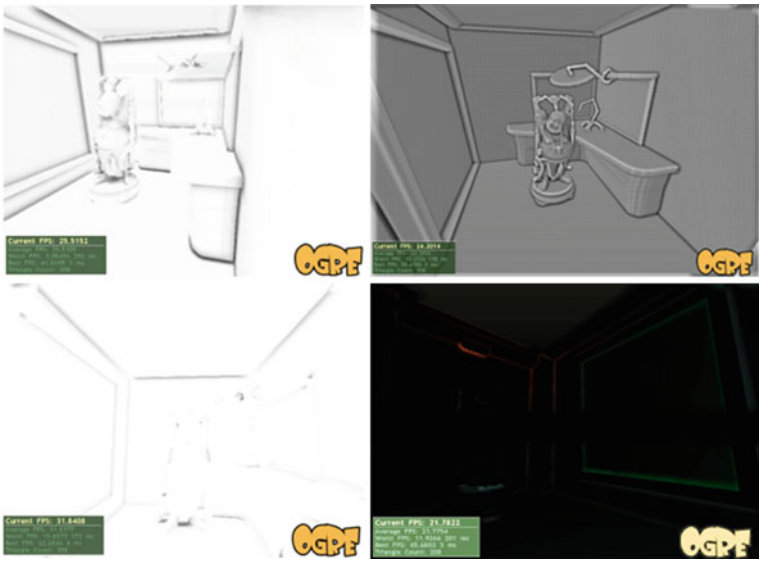


Fig. 20 Screen space ambient occlusion: classic, Crytek, volumetric, and volumetric with color bleeding

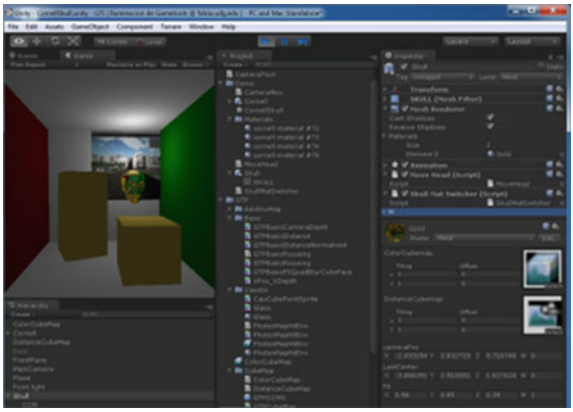


Fig. 21 GameTools illumination effects in Unity3D

4 Use Cases

The updated GameTools libraries have been used in two serious games created with the Unity3D engine. These games have been designed in collaboration with the faculty of arts and the faculty of education of the University of Girona, and they are to be used to help teaching local history to middle school students. An agile software development methodology has been followed to create these games.

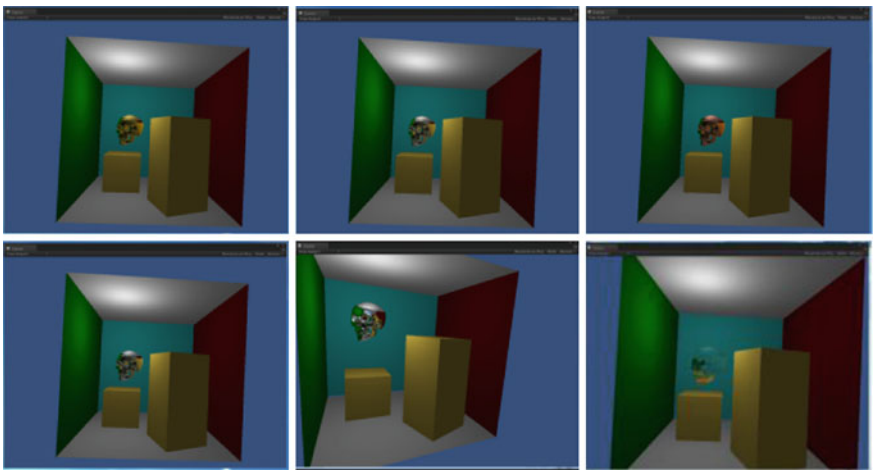


Fig. 22 Realistic gold, silver, copper, and aluminum metals, ideal metal and glass in the Unity3D engine (Windows)

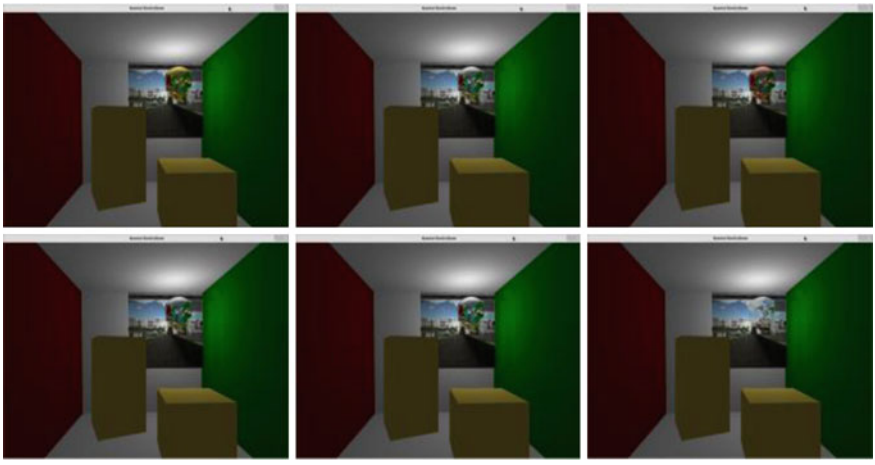


Fig. 23 Realistic gold, silver, copper, and aluminum metals; ideal metal and glass in the Unity3D engine (Linux + Wine)

Preliminary testing with end users has been carried out, and more widespread tests are planned for the near future.

The first game is called Jaume I and simulates the conquest of Mallorca by James I the Conqueror in 1229. The second game is legends of Girona. The legends of Girona game explores the different legends concerning the city of Girona. In particular, the legend of the miracle of Saint Narcis, who repelled the siege of Girona in 1285, has been implemented. The objective of these two games

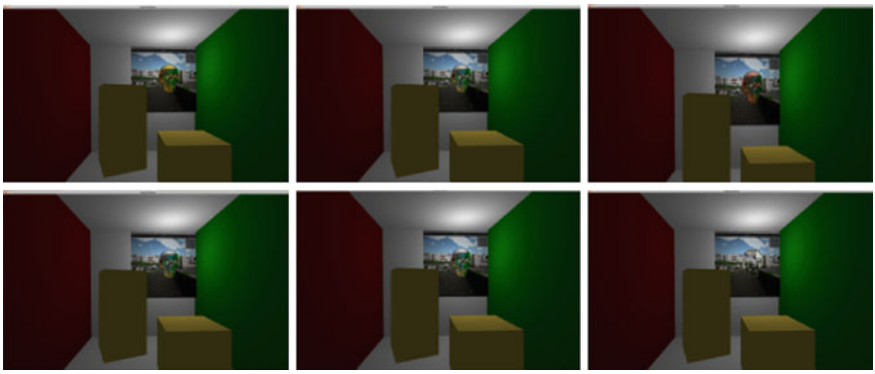


Fig. 24 Realistic gold, silver, copper, and aluminum metals in the Unity3D engine (Mac OSX + Wine)

is to provide an enjoyable way to study the history of the kingdom of Aragon in the middle ages (Jaume I) and specifically the history of the city of Girona (Legends).

The third use case is not strictly a serious game, but a computer version of a board game called Nine Men's Morris (and its more common variants). However, since this game has been played since roman times and was very popular in medieval times across most of Europe, it is of interest for interactive museum exhibits dealing with roman or medieval times. Because of this, we may consider it a serious game in the context of cultural heritage. The objective of this game is to familiarize museum visitors with the different variants of this milenary board game, and the materials commonly used during roman and medieval times. This game has been designed using a classical iterative development model, and informal testing has been carried out during the development of the game.

No changes in the design of the games were required to integrate the updated GameTools effects. Drag-and-drop was used to indicate the position of the CLOD trees and the realistic materials.

4.1 Jaume I: The Battle of Portopí

Jaume I, a strategy game, reproduces a historical battle in thirteenth century, with archery, infantry and cavalry, won by king of Aragon Jaume I which led him to conquer Mallorca. The game has been developed in collaboration with the faculty of arts of the University of Girona, and tries to reproduce the historical characters and skirmish situations. Successful players will develop mastery of the history of the Aragon kingdom before unification with the Castilian kingdom to form Spain.

Figure 25 shows a continuous level of detail tree created using the GTGeometry libraries at the beginning of the game.

Fig. 25 Screenshots of the Jaume I serious game, including a procedural tree from GTGeometry at different LODs

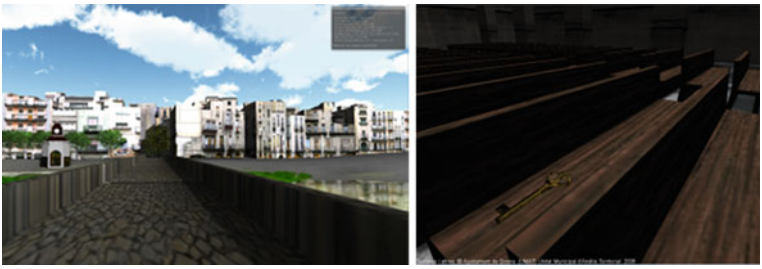


Fig. 26 Screenshots of the Legends of Girona serious game, including a procedural tree from GTGeometry and a realistic gold key from GTIllumination, running on Mac OSX and Windows

4.2 Legends of Girona

Legends of Girona is a first/third person educational game developed together in collaboration with the faculty of education of the University of Girona. The game simulates a legendary happening in the town during a foreign invasion in the thirteenth century. The successful player develops mastery of the history of the town of Girona and its ancient urban and geographical disposition through navigating and exploring game artifacts without being captured. Eventually the player needs to find a way out of a labyrinth in an old church crypt.

Figure 26 (left) shows a screenshot of the game with a tree showing simplified geometry. When the user crosses the bridge, the tree geometry is refined smoothly in real time using the continuous level of detail routines from our libraries. Figure 26 (right) shows another screenshot, which displays a gold key with realistic reflections in real time.

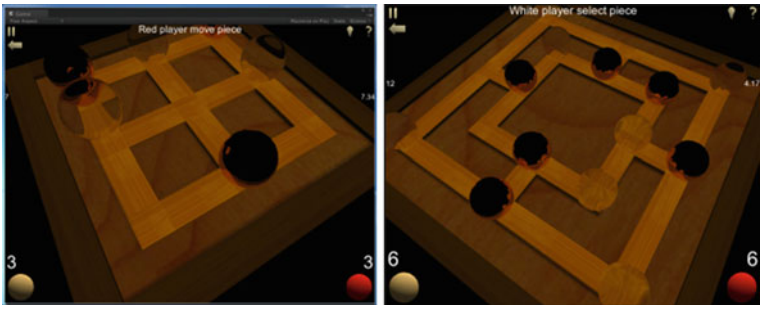


Fig. 27 Screenshots of the Nine Men's Morris game, including realistic cooper and glass materials from GTIllumination, running in Windows15.5 conclusions and future work

4.3 Nine Men's Morris

Nine Men's Morris is a traditional board game similar to Tic-Tac-Toe. While not strictly a serious game, we are interested in including this game (and its many variants) in interactive museum displays in the United Kingdom and Northern Europe, because this game was very popular there in medieval times.

We have implemented different variants of the Nine Men's Morris strategy game, and included an artificial intelligence opponent. We have used the realistic copper and glass materials from the GTIllumination libraries to distinguish the two players, and a wood material for the board. These materials have been used since ancient times for board games, and we expect the increased realism provided by our routines to help set the scenario in museum exhibits. Figure 27 shows a screenshot of the game; notice the multiple inter-reflections and refractions visible in the marbles' surfaces.

5 Conclusions and future work

We have presented in this paper a description of the updated GameTools libraries, showing how they can be used to create advanced graphic effects in popular game engines. Users of the routines are not expected to have in-depth knowledge of graphic algorithms, and this makes the routines especially suited for inclusion in serious games created by programmers with backgrounds other than computer graphics. The main advantage of the routines is that they can be used to increase the performance or realism of serious games without requiring changes in the design of the games. Three use cases are provided: the games Jaume I, Legends of Girona and Nine Men's Morris.

Our plans for future work include adding a visibility culling module to our routines, adding more game engines and platforms (including popular consoles

such as Microsoft Xbox, Nintendo Wii, and Sony PlayStation), and adding more advanced effects (other simplification and LOD techniques, more realistic shadows, other screen space ambient occlusion, and/or global illumination algorithms), in collaboration with academia.

Acknowledgments This work has been supported by the research projects coded TIN2010-21089-C03-01 and IPT-2011-0885-430000 (Spanish Commission for Science and Technology) and by grants VALTEC09-2-0118 and 2009SGR643 (Catalan Government).

References

ArUco (2012) ArUco: a minimal library for augmented reality applications based on OpenCv. http://www.uco.es/investiga/grupos/ava/node/26. Accessed 23 January 2012

Bittner J, Wimmer M, Piringer H, Purgathofer W (2004) Coherent hierarchical culling: hardware occlusion queries made useful. url http://www.cg.tuwien.ac.at/research/publications/2004/Bittner-2004-CHC/. Proceedings of the Eurographics 2004. Grenoble, France

Domingo A, Escriva M, Abad F Lluch J, Camahort E, Vivo R (2007) Continuous LODs and adaptive frame-rate control for spherical light fields. In: Geometric modeling and imaging–new trends pp 73–78 doi:http://doi.ieeecomputersociety.org/10.1109/GMAI.2007.14

Fast light toolkit (FLTK) (2012) http://www.fltk.org/. Accessed 23 January 2012

GameTools (2008a) GameTools project newsletter. http://www.gametools.org/downloads/GTP_newsletter3.pdf. Accessed 20 January 2012

GameTools (2008b) GTP. http://www.gametools.org/. Accessed 20 January 2012

González C, Castelló P, Chover M (2007a) A texture-based metric extensión for simplification methods. In: Braz J, V′azquez PP, Pereira JM (eds) GRAPP (GM/R), Institute for systems and technologies of information, control and communication (INSTICC). Barcelona, Spain pp 69–76

González C, Gumbau J, Chover M, Castell P (2007b) Simplificación de mallas para juegos, vol 1. In: Congreso Español de Informática Gráfica (CEIG 2007), Eurographics Association, Zaragoza, Spain

Gumbau J, Ripolles O, Chover M (2007) Lodmanager: a framework for rendering multiresolution models in real-time applications. Proceedings of the WSCG'2007 short communications papers. Plzen, Czech Republic pp 39–46

Hadwiger M, Varga A (1999) Visibility culling. In: Proseminar Wissenschaftliches Arbeiten, 1998–1999 url: http://www.cg.tuwien.ac.at/∼msh/viscull.pdf

Lazányi I, Szirmay-Kalos L (2005) Fresnel term approximations for metals. Proceedings of the WSCG 2005, short papers. Plzen, Czech Republic pp 77–80

Lazányi I, Szirmay-Kalos L (2006) Indirect diffuse and glossy illumination on the GPU. In: Engel W (ed) ShaderX5: advanced rendering techniques. Charles river media, Chilton pp 345–358

Méndez A, Sbert M, Cata J, Sunyer N, Funtane S (2005) Real-time obscurances with color bleeding. In: Engel W (ed) ShaderX4: advanced rendering techniques. Charles river media, Chilton

OGRE (2012) OGRE—Open Source 3D Graphics Engine. http://www.ogre3d.org/. Accessed 20 January 2012

OpenCV (2012) Welcome–OpenCV wiki. http://opencv.willowgarage.com/wiki/. Accessed 23 January 2012

Rebollo C, Gumbau J, Ripolles O, Chover M, Remolar I (2007a) Fast rendering of leaves. Proceedings of the 9th international conference on computer graphics and imaging (IASTED) CGIM 2007. ACTA Press, Anaheim, pp 46–53 url http://dl.acm.org/citation.cfm?id=1710707.1710717

Rebollo C, Remolar I, Chover M, Gumbau J, Ripolles O (2007b) A clustering framework for real-time rendering of tree foliage. J Comp 2(4):57–67

Seward J, Nethercote N (2005) Using valgrind to detect undefined value errors with bit-precision. Proceedings of the annual conference on USENIX annual technical conference. USENIX Association, Berkeley, p 2–2 url http://dl.acm.org/citation.cfm?id=1247360.1247362

Shark3D (2012) Shark 3DTM by Spinor GmbH. http://www.spinor.com/. Accessed 20 January 2012

Szécsi L, Szirmay-Kalos L, Sbert M (2006) Light animation with precomputed light paths on the GPU. Proceedings of graphics interface 2006 (GI2006), Canadian Information Processing Society, Toronto pp 187–194

Szirmay-Kalos L, Aszódi B, Lazányi (2005a) Ray-tracing effects without tracing rays. In: Engel W (ed) ShaderX4: lighting and rendering. Charles River Media, Chilton

Szirmay-Kalos L, Aszódi B, Lazányi I, Mátyás P (2005b) Approximate raytracing on the GPU with distance impostors. Compu Grap Forum 24(3):695–704

Szirmay-Kalos L, Lazányi I (2006) Indirect diffuse and glossy illumination on the GPU. Proceedings of the SCCG 2006. Smolenice castle, Slovakia pp 29–35

Tóth B, Szirmay-Kalos L (2007) Fast filtering and tone mapping using importance sampling. Proceedings of the WSCG 2007 short papers. Plzen, Czech Republic pp 47–52

Umenhoffer T, Szirmay-Kalos L (2005) Real-time rendering of cloudy natural phenomena with hierarchical depth impostors. In: Eurographics short paper. Dublin, Ireland pp 65–68

Umenhoffer T, Szirmay-Kalos L, Szíjártó G (2006) Spherical billboards andtheir application for rendering volumetric data. In: Engel W (ed) ShaderX5: advanced rendering techniques. Charles River Media, Chilton pp 275–286

Umenhoffer T, Szirmay-Kalos L (2007) Robust diffuse final gathering on the GPU. Proceedings of the WSCG 2007 full papers. Plzen, Czech Republic pp 121–128

Umenhoffer T, Szirmay-Kalos L (2006) Spherical billboards for rendering Explosions.In: Graphics Interface. Quebec, Canada pp 57–64

Umenhoffer T, Patow G, Szirmay-Kalos L (2007) Robust multiple specular reflections and refractions. In: Nguyen H (ed) GPU Gems 3. Addison, Wesley

Unity (2012) Unity: game development tool. http://unity3d.com/unity/. Accessed 20 January 2012

Valgrind (2012) Valgrind home. http://valgrind.org/. Accessed 23 January 2012

Wine (2012) WineHQ—Run Windows applications on Linux, BSD, Solaris and Mac OS X. http://www.winehq.org/. Accessed 29 February 2012

Xfrog (2012) Xfrog—news. http://xfrog.com/. Accessed 23 January 2012

Pink Dolphins: A Serious Simulation Game

Noel K. H. Chia, Norman K. N. Kee, Yiyu Cai and Nadia Thalmann

Abstract Creating a 3D virtual dolphin requires more than just virtual reality (VR) technological know-how. It also involves a good understanding of the behavior and habits of a particular species of dolphin, and in this case, the chosen species of Delphinidoe is the Indo-Pacific humpback dolphin (known by its scientific name as *Sousa Chinensis*), also known by other names, such as Chinese white dolphin and Pacific humpback dolphin. In Singapore, they are commonly known as pink dolphins because of their color as they mature into adulthood. This chapter describes how to model a virtual pink dolphin resembling closely to the real one basing on a good understanding of the pink dolphin in terms of its size, behavior and habits, and its acrobatic acts. The virtual pink dolphins model are used for a serious game developed to assist the special needs education.

Keywords Dolphin · Virtual reality · Simulation and modeling · Game

1 Introduction

Dolphins are generally popular with humans because they appear friendly, intelligent, are graceful, and their performance while jumping in the ocean fascinates humans. The family Delphinidae (dolphins) evolved about 10 million years ago

N. K. H. Chia · N. K. N. Kee
National Institute of Education, Nanyang Technological University, Singapore, Singapore
e-mail: khchia@nie.edu.sg

N. K. N. Kee
e-mail: knkee@nie.edu.sg

Y. Cai (✉) · N. Thalmann
Nanyang Technological University, Singapore, Singapore
e-mail: myycai@ntu.edu.sg

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_13,

during the Miocene Period. Their early ancestors became fully aquatic around 38 million years ago. The *Basilosaurus* and *Dorudon* were two such aquatic ancestors of dolphins. In fact, they resembled like modern-day dolphins and whales. These ancient whales or dolphins were known as Archaeoceti and they probably evolved in the early Paleocene Period, around 65 million years ago. They were completely land-based and it was only in the following Eocene Period that they began living in water more than on land. One such terrestrial ancestor of dolphins is known as the Pakicetus that resembled the modern-day wolf (Fordyce 1980; Fordyce and Barnes 1994).

Today, there are as many as 17 genera with about 40 different species of dolphins in them. Dolphins are close cousins of whales and porpoises. They are warm-blooded mammals. These toothed whales belong to the order of *Cetacea* and the family of *Delphinidae*. They like to live in warm, shallow waters along the coast. They can also be found in open seas and oceans as well as in guls, lagoons, bays, and even harbors. Sometimes they come into large rivers, and sometimes they go further out to sea. In fact, they are found in almost every part of the world, with the bottlenose dolphin being the most widely distributed. Dolphins live in family groups called pods. There are often about 12 dolphins in a pod, but sometimes several hundred dolphins get together. They do not usually migrate, but they may follow schools of fish that are moving. These fish provide them the source of food.

As there are so many different species of dolphins, the authors have chosen to focus on a particular species of dolphins known as the Indo-Pacific humpback dolphins (*Sousa Chinensis*) which they have been researching on for the last 5 years since 2007 in partnership with the Underwater World Singapore based in Sentosa—a holiday resort island south of Singapore mainland.

This chapter is organized as follows: Section 2 examines the species of Indo-Pacific humpback dolphins covering on the anatomy of a dolphin, functions of some selected parts, its behavioral pattern. Section 3 discusses the technical aspects that were taken into consideration in the design of a virtual pink dolphin game to assist the special needs education. Section 4 concludes this research.

2 Dolphin Anatomy, and Behavioral Patterns

Humpback dolphins belong to the genus of *Sousa* (Fig. 1) and they are characterized by the conspicuous humps ahead of the dorsal fin, which is to some degree falcate, as well as a careen on a ventral side Their pectoral flippers are considerably small and the tail flukes have a well-defined median notch. On each side of the jaw there are 30–34 small coned-shaped teeth. Their pectoral flippers are considerably small and the tail flukes have a well-defined median notch. On each side of the jaw there are 30–34 small coned-shaped teeth (Ross 2002; Rice 1998).

Fig. 1 An Indo-Pacific humpback dolphin

Different dolphin species vary in size, lying anywhere between 1.2 and 9.5 m in length and between 40 kg and 10 tons in weight. An adult humpback dolphin can reach from 1.8 to 2.4 m and weigh in the range of 100–139 kg.

Newborn calves are a cream or pearl shade of white, whereas the adults have a more dull off-white coloring from the tail to the snout. Their flanks are somewhat of a dark gray, and their belly is a lighter shade of gray. These humpback dolphins can be found swimming close to shore along the West African coast as well as along the coast of the Indian Ocean from South Africa to Australia. The main diet of the humpback dolphins consists of mullet and other fish, though the feeding habits are widely unknown, as this animal is not widely known itself.

The taxonomy of the *Sousa* genus is complex and continues to be the topic of debate among the experts. As many as 5 species have been proposed and they are *S. chinensis* (Chinese white dolphin, Pacific humpback dolphin, or Indo-Pacific humpback dolphin), *S. plumbea* (Indian humpback dolphin or Plumbeous humpback dolphin), *S. teuszi* (Atlantic humpback dolphin), *S. Lentiginosa, and S. Borneensis* (Ross 2002). However, Rice (1998) argued that there are only three sub-species of *Sousa*, viewing the Indo-Pacific as an intermixing between two other sub-species named simply the Indian and the Pacific.

2.1 Anatomy of a Dolphin

In designing a virtual dolphin, it is important to have the correct size so that once it is created, it should not be under or over exaggerated, but be true or close to its real proportions. From Fig. 2, beginning with the mouth known as beak or rostrum in

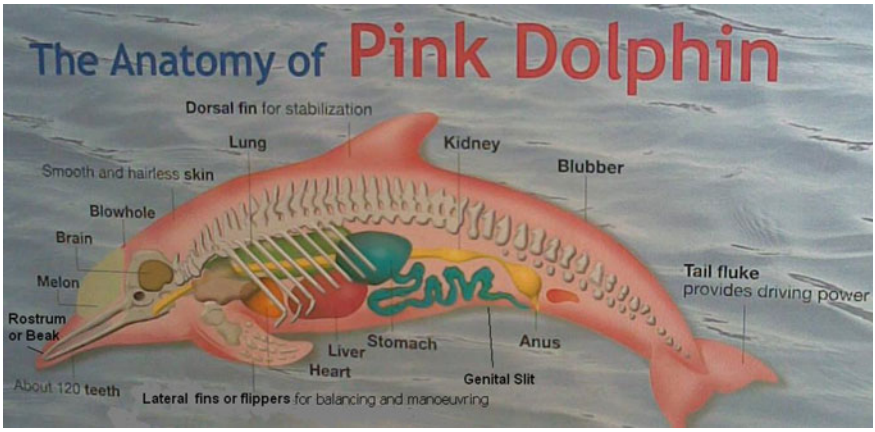


Fig. 2 Parts of a dolphin (Photo Courtesy of Underwater World Singapore)

which the lower jaw is longer than its upper. The teeth of a dolphin are conical, arranged around its beak-shaped mouth. It has a wide forehead known as melon. Further up the melon is a blowhole, which the dolphin uses to get air. Dolphins are mammals, not fish, and so they do not have gills but lungs. On the back of its body is the dorsal fin. Further down its body is the tail with two lobes known as flukes. On both sides of the body are the pectoral fins of flippers. It navel (the dolphin's belly button is found close to its genital slit which conceals the reproductive organs and anus.

2.2 Dorsal Bursa and Dolphin Whistles

Dolphins can be very noisy animals and they make three main types of nasal vocalizations—whistles, burst-pulse sounds, and clicks—from the air-sacs below their blowholes.

Unlike human, a dolphin does not have vocal cords in its larynx to make sounds. Anatomical and bioacoustic research suggests that a dolphin has two tissue complexes in the nasal region called dorsal bursa, which includes what are known as phonic lips (structures that project into the nasal passages) is probably the most likely site. The phonic lips can work independently or simultaneously of each other. As air pushes through the nasal passage passing by the phonic lips, the surrounding tissue vibrates to produce one or more types of sounds.

The sounds that dolphins produce resemble grunts, moans, squeaks, and trills that can be made any time and at considerable depths varying in volume, wavelength, frequency and pattern. The frequency of dolphin sounds ranges from 0.2 to 150 kHz. The lower frequency vocalizations (about 0.2–50 kHz) are likely used in social communication between and among dolphins. Social signals have their most

energy at frequencies less than 40 kHz. Higher frequency clicks (40–150 kHz) are primarily used for echolocation.

Chia and Kee (2012) investigated sounds made by bottlenose dolphins to see the response of children with autism. Their study suggested that these children preferred buzzing clicking, which they felt more ambient, than the other three sounds: clicking, creaking, and squeaking. It would be interesting to create a dolphin avatar or virtual dolphin that could make buzzing clicks to give a calming effect to those interacting with it. Of course, to create a virtual dolphin that resembles a real one, it must be able to make all the four sounds.

Dolphins rely heavily on echolocation for two reasons. First, they live in an underwater habitat that has favorable acoustic characteristics. Second, in such an increasing depth underwater, vision is extremely limited in range due to absorption, i.e., sunlight is absorbed and the amount of visible light diminishes. Because absorption is greater for the long wavelengths than for the short ones, the color spectrum (as in the rainbow) is rapidly altered with increasing depth. For instance, objects that are white at the surface appear bluish underwater. Similarly, red objects will appear dark or even black. Although light penetration will be less if water is turbid, in the open ocean with very clear water, <25 % of the surface light will reach a depth of 10 m. At a depth of 100 m, the light present from the sun is normally about 0.5 % of that at the surface. Such details must be taken into consideration when designing the virtual aquatic habitat for the virtual dolphins.

Dolphins emit a focused beam of high-frequency clicks in the direction that their head is pointing. These sounds are generated by passing air from the bony nares through the phonic lips (Cranford 2000), and are reflected by the dense concave bone of the dolphin's cranium and an air sac at its base. On the dolphin's head is a large fatty organ called melon that helps to modulate the focused beam (Ketten 1992, 2000). It is composed of lipids of differing densities and serves as an acoustic (Ketten 1992).

Most dolphins use clicks in a series, or click train, for echolocation to find their location, identify objects as well as their sizes and see how far they are from them as well as the location of other animals. Different click train production rates give rise to the familiar barks, squeals, and growls of the dolphins. A click train with a repetition rate more than 600 clicks per second is called a burst pulse (Ketten 1992). In the species of bottlenose dolphins, their auditory brain response resolves individual clicks up to 600 clicks per second, but yields a graded response for higher repetition rates (Ketten 1992, 2000).

2.3 The Behavioral Patterns of Dolphins

Dolphins are well adapted to living underwater. A dolphin possesses sensitive skin on its lower jaw which allows it to identify small objects, and it has a blowhole on top of its head that enables it to breathe air from the surface.

The streamlined body of a dolphin has few appendages (like ears, arms, and legs) sticking out to slow it down as it swims through the water. Its sleek body with smooth rubbery skin which secretes an oily substance that enables it to swim through water smoothly and speedily (Zagzebski 2009; Fish 2006).

However, the dolphin takes more than just a sleek body to maneuver in the water. When dolphins are going after their food, they not only swim at the surface, they also dive to great depths. They need rudders and propellers and these are their dorsal fin (i.e., the fin that sticks up from its back) which acts as a stabilizer or a rudder. The dorsal fin has no bones in it, unlike the fins on fish. Instead, it is made up of dense tissue, somewhat like a ridge of thick, folded skin (Zagzebski 2009). The pectoral flippers on the sides of the dolphin are used for steering, balancing, slowing down, and stopping. They are not used for moving forward through the water (Zagzebski 2009). Forward motion is created by the dolphin's tail known as fluke, which is not perpendicular (up and down) like a fish or a shark. Instead, its fluke spreads out side to side. When a dolphin swims, it pushes its fluke up and down to help it dive or leap out of the water. When swimming at high speed, the dolphin leaps out of water. Leaping helps the dolphin to conserve its energy as it can move a longer distance with one long jump compared to swimming in denser media of water relative to air (Fig. 3) (Au and Weihs 1980).

In other words, the flukes serve as the propeller. Powerful muscles running along the backbone and sides of the dolphin's body move the tail up and down in the water, providing the power that pushes the animal through the water, or deep into the ocean (Zagzebski 2009). In fact, one way to remember that dolphins, like whales, are not fish which move their bodies sideways when they swim; dolphins move their bodies up and down.

In creating a virtual dolphin to swim like one, its motions in the water must be observed carefully, especially how it actually executes its various strokes and maneuvers using its dorsal fin, its pair of pectoral flippers and its tail fluke.

Fig. 3 Leaping out of water

3 From Physical Pink Dolphins to Virtual Pink Dolphins

In this research, virtual pink dolphins are modeled following the real dolphins with the Underwater World Singapore. They have different sizes, different characters, and different behavioral patterns. In the followings, we detail the modeling, visualization, interaction, user interface and simulation with the virtual pink dolphins serious game (Fig. 4).

3.1 Dolphin Shape Modeling

We model pink dolphins based on their anatomy. Efforts are made to model the shapes, skeleton, and skins of pink dolphins. Geometric representation is used to create smooth surface of virtual pink dolphins which is then tessellated into triangular meshes for digital representation. Digital geometry processing techniques are further applied to produce compact and optimal representation of the virtual pink dolphins (Fig. 5). Based on the dolphin shapes created, we insert the skeleton in the dolphin surface model for animation and simulation. The virtual pink dolphins such modeled have typical dynamic behaviors such as swimming, jumping, whistling, etc., which serve a basic role in communication among dolphins. Figure 6 shows three virtual pink dolphins modeled following their real counterparts from a dolphin family in UWS (Left: the daughter, Middle: the son, and Right: the mother).

3.2 Dolphins Visualization

The 3D modeling of the virtual pink dolphins can be easily used for realistic visualization in either stereographic or normal modes. For stereographic visualization, active shutter glasses or anaglyph glasses can be used for viewers to have 3D perception with the virtual pink dolphins.

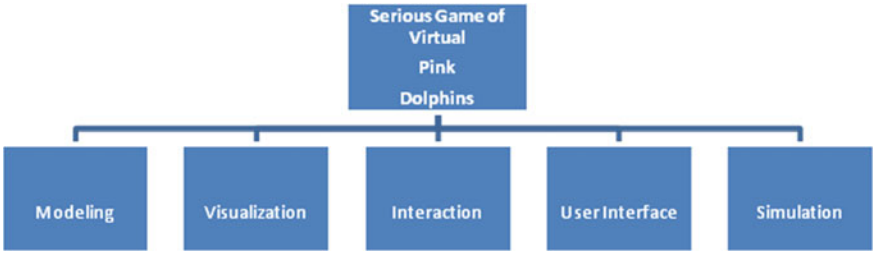


Fig. 4 Overall structure of the virtual pink dolphin system

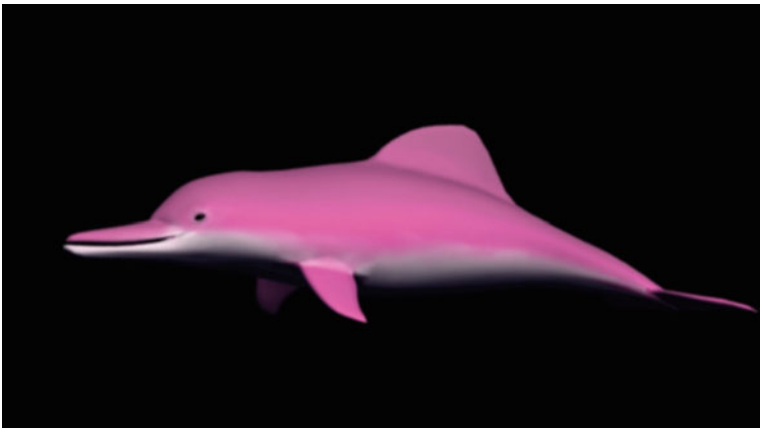


Fig. 5 A virtual dolphin modeled with a multilayer structure

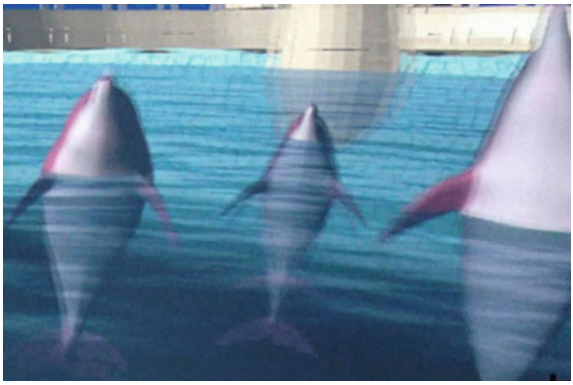


Fig. 6 Three virtual dolphins modeled following their real counterparts from a dolphin family in UWS

3.3 Real-Time Interaction

The virtual pink dolphins also have certain interactive functions. For instance, they can be interactively controlled to swim, leap, and whistle. For this, both geometrically and physically based modeling techniques are implemented with the virtual pink dolphins. One of the most important physical modeling tasks is to deal with collision detection. We use Graphic Processing Unit (GPU) technology to speed up the physical modeling process. The interactive control, however, will be carried out through the user interface specially designed.

3.4 Natural User-Interface

Instead of using mouse and keyboard to control the virtual dolphin interaction, we implement an easy-to-use natural interface to control the virtual pink dolphins with hand gesture. The Microsoft Kinect device is integrated with the virtual pink dolphin system through its natural interface programming.

3.5 Serious and Simulated Game of Virtual Pink Dolphins

We have developed an interactive simulation game with virtual pink dolphins for children with autism or other special needs. A 3D immersive environment is designed and developed allowing those kids to interact with the virtual pink dolphins. They can use their hand gestures to control the dolphins' behaviors (swimming, leaping, whistling, etc.). We hope this interactive game can help those children improving their nonverbal communication skill. Figure 7 shows two children from a special needs school in Singapore playing the interactive game.

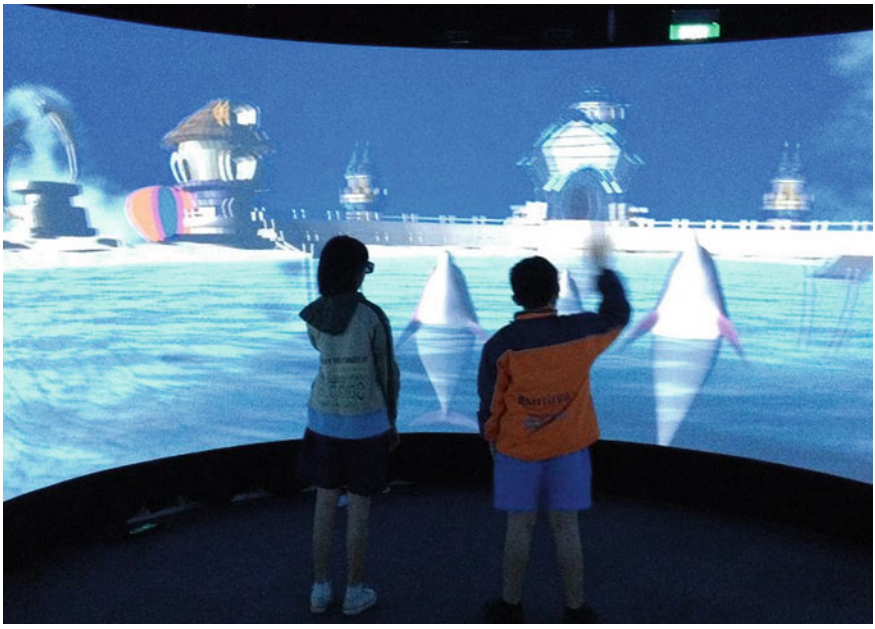


Fig. 7 Two special needs children are interacting with the virtual pink dolphins

Fig. 8 A reflection by a special need child after her interactive play with the virtual pink dolphins

4 Conclusion

The Virtual Pink Dolphins Project is a new initiative to study the VR technology for applications of special needs education. A serious game is developed with a focus on improving the nonverbal communication skills of special needs children. For this aim, significant research has been carried out to understand the anatomy of the pink dolphins, and their behavioral patterns. From there, techniques of 3D modeling, visualization, interaction, natural interface, and simulation are investigated for the development of the virtual pink dolphin game.

Five sessions have been conducted for a dozen of children from a local special needs school. Special teaching plan and teaching activities are designed based on the dolphin context. These children learn anatomy, vocabulary, and shapes of the pink dolphins. They also learn hand gesture-based nonverbal communication through interactive playing with the virtual pink dolphins. Figure 8 is one of the reflections by a child after her game play with the virtual pink dolphins.

There are lots of work can be done with the virtual pink dolphins. A pilot study on the use of virtual pink dolphins to assist children with autism was recently reported (Cai et al. 2013). We believe edutainment concept and edutainment technology (Cai et al. 2006) have big room to develop for the special needs education.

Acknowledgments The project team would also to thank the support from the Underwater World Singapore. This project is sponsored by Temasek Trust Funded Singapore Millennium Foundation and NTU's Institute for Media Innovation under a Seed Grant Scheme.

References

Au D, Weihs D (1980) At high speeds, dolphins save energy by leaping. Nature 284(5756):548–550

Cai YY, Lu BF, Fan ZW, Indhumathi C, Lim KT, Chan CW, Jiang Y, Li L (2006) Bio edutainment: learning life science through X gaming. Comput Graph 30(2006):3–9

Cai Y, Chia N, Thalmann D, Kee N, Zheng J, Thalmann N (2013) Design and development of a virtual dolphinarium for children with autism. IEEE Trans Neural Syst Rehabil 21(2):208–217

Chia NKH, Kee NKN (2012) A study on the responsivity of children with autism to dolphin sounds during manipulative activities. Paper presented at the meeting of the association of educational therapists-Singapore, Chapter Study Group, Singapore, 8 June 2012

Cranford TW (2000) In search of impulse sound sources in odontocetes. In: Au WWL, Popper AN, Fay RR (eds) Hearing by whales and dolphins. Springer-Verlag, New York, pp 109–156

Fish FE (2006) The myth and reality of Gray's paradox: implication of dolphin drag reduction for technology. Bioinspiration Biomimetics 1:17–25

Fordyce RE (1980) Whale evolution and oligocene southern-ocean environments. Palaeogeogr Palaeoclimatol Palaeoecol 31:319–336

Fordyce RE, Barnes LG (1994) The evolutionary history of whales and dolphins. Annu Rev Earth Planet Sci 22(1):419–455

Ketten DR (1992) The marine mammal ear: specializations for aquatic audition and echolocation. In: Webster D, Fay R, Popper A (eds) The evolutionary biology of hearing. Springer-Verlag, New York, pp 717–750

Ketten DR (2000) Cetacean ears. In: Au W, Fay R, Popper A (eds) Hearing by whales and dolphins. Springer-Verlag, New York, pp 43–108

Rice DW (1998) Marine mammals of the world: systematics and distribution. The Society for Marine Mammalogy, San Francisco Special Publication 4

Ross GJB (2002) Humpback dolphins: *Sousa chinensis*, *S. plumbea* and *S. teuszi*. In: Perrin WF, Wursig B, Thewissen JGM (eds) Encyclopedia of marine mammals. Academic Press, San Diego, pp 585–589

Zagzebski K (2009) How do whales and dolphins swim? National Marine Life Center, Buzzards Bay

Gender Differences in Learning Molecular Biology Using Virtual Learning Environments

Sandra Tan

Abstract In another project of ours, an experimental study quantified the effects of using an immersive three-dimensional virtual reality to visualise molecular biology phenomenon. Results indicated significant increases in attitudes and behaviours towards molecular biology in all participants. However, major improvement in achievement in molecular biology was observed only in male students, while marginal improvements in achievement by female students were not significant. The gap in achievement between genders, despite similar improvements in motivation, suggests that the male and female students imbibe and process information differently. Forty-six participants were thus followed up with qualitative interviews to gain insights into how male and female students learn molecular biology. Qualitative surveys suggest that, prior to exposure to virtual reality-based visualisations, both male and female students relied heavily on rote memorisation to do well in tests. This tendency dissipated in male students as they moved towards, perhaps as a result of the virtual reality simulations appealing to their preferences for a "bigger picture", helping them make connections between different structures and processes to enhance understanding. However, female students continued to rely on memorising the prescribed syllabus to help themselves in examinations. It seems that their tendency for the written word, and to work reflectively predisposes them towards memorisation as a learning strategy. Examination of learning behaviours using the VARK framework shows that most male students prefer learning by multiple modalities while female students exhibited an overwhelming preference for reading and writing. The VARK findings were congruent with male and female learning behaviours tracked by the Learning Styles Index, where males were sensory, visual and verbal in acquiring information and active in their learning and global in their learning perspectives. Conversely, females tended to be intuitive, and reflective in learning, preferring sequential information dissemination. Because lessons using virtual

S. Tan (✉)
Hwa Chong Institution, Singapore, Singapore
e-mail: sandra@hci.edu.sg

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_14,

reality simulations also involve questioning, discussion and model construction, these activities clearly catered to the preferences of male students. One of the original objectives of the virtual reality simulations was to show students the interconnectedness of sub-cellular processes and structures; the information presented was not sequential, but served to demonstrate the synergism in different parts of the cell. This global perspective would again appeal more to the learning preferences of male students. While female students may find such activities refreshing—hence driving motivation and engagement scores—these contributed little to their actual learning. Therefore, more efforts are needed to help female students on learning structure information and making connections across different topics in molecular biology. More broadly, the technology used in mixed gender classrooms and differentiated instruction programmes need to be carefully considered.

Keywords Virtual reality · Molecular biology · Learning behaviours · Gender differences

1 Introduction

The motivation for the present study stems from observations of how male and female students grappled with understanding molecular biology. In a previous quantitative study (Tan and Waugh 2013), participants were randomly divided into two groups. Students in the experimental group were taught using the virtual reality module to enhance visualisation of molecular biology phenomenon, while students in the control group were taught the same content in the traditional way. Before and after the experiment, students in both groups did a survey on attitude and behaviour towards molecular biology and a molecular biology achievement test. Data were then analysed using Rasch measurement models (Andrich 1988; Andrich et al. 2010). Motivation increased in all participants in the experimental group at posttest. However, the intervention for virtual reality-based visualisation in molecular biology produced significant improvement in terms of molecular biology achievement only in male participants (Tan and Waugh 2013). It was perplexing that females were not benefiting from the intervention, particularly when juxtaposed against the enhanced motivation scores in both groups. Hence, this chapter will present our study with a focus on the learning behaviours of secondary school male and female students in molecular biology.

Individual differences in learning and achievement have persistently challenged educators. During the late 1960s, Cronbach and Snow (1969) postulated that individualizing teaching to the needs of learners would improve satisfaction and achievement. Their Aptitude Treatment Interaction Theory (ATI) argued that teaching methods differentially affect students' learning because teaching methods place varying demands on learners and account for learners' achievement.

Interpolations from the ATI theory have resulted in multiple ways of examining teaching and learning from left and right brain emphasis (Bancroft 1995), cognitive styles (Messick 1984), learning strategies (Butler 1994), to personality styles (Cooper and Miller 1991).

The purpose of the current investigation is to determine whether males and females differ with reference to their learning behaviours in a molecular biology classroom. Several studies on individual learning differences argue that essential differences exist between male and female learning styles. For example, Dwyer (1998) reported that trait/context communication apprehension significantly relates to learning style preferences for females. Dwyer' study recommended removing hierarchical classroom barriers in terms of classroom organization or forms of threats that have the potential of invalidating females' learning efficacy, independent of course content. Lundeberg et al. (1994) also found that there were significant differences in the confidence of undergraduate males and females to test-item responses. Mann (1994) examined the development of women and girls within a hierarchical power structure relative to the development of their self-esteem, and general academic performance in certain subjects. Faced with several challenging learning conditions such as instructor bias and passivity, and institutions that destroy friendship networks, Mann (1994) found that females were less likely to exemplify their learning styles in subject areas such as math and sciences. Therefore, he encouraged teaching techniques that place more emphasis on collaboration and hands-on learning. Such reports, however, are drawn from student profiles in Europe and the United States. Moreover, these do not consider the influence of technology in interaction and learning, nor take into account how male and female students might respond to a relatively new subject area in biology. It is hoped that insights gained here would help translate into more effective pedagogies in technology-intensive classrooms in Singapore.

2 Implementation and Evaluation of Study

Twenty-three male participants and 23 female participants were interviewed separately about their perceptions of, and learning behaviours in molecular biology, before and after exposure to immersive virtual reality simulations detailing sub-cellular structures and processes. Specifically, the participants were asked how they studied for a molecular biology exam. In particular, they were asked to list specific learning behaviours, and to explain why these were effective for them. Responses were carefully recorded, and key words or categories representing the raw responses were identified. The responses were then classified into different learning preferences and behaviours using the VARK framework (Fleming and Mills 1992) and Felder and Silverman's Index of Learning Styles (1988).

3 Results and Discussions

3.1 Female Students Rely Heavily on Memorisation Even with Virtual Reality Supported Learning

A predominant discourse amongst participants interviewed was the heavy reliance on rote memorisation as the means to score well in tests. Prior to learning molecular biology via virtual reality simulations, all 46 participants interviewed reported difficulty grappling with molecular biology and believed that memorisation was necessary; even students of higher ability regarded it as an effective way to do well.

This finding is corroborated by data on students' perception of item difficulties in the molecular biology achievement test. In general, students found items testing specific knowledge easier than those testing more general concepts. This indicates a propensity to memorise specific common answers in biology and an inability to relate these to general theory. The fact that memorization was a key examination strategy for molecular biology indicates a need to re-examine teaching and learning in schools and indicate that more resources need to be developed to help students understand molecular biology concepts.

After exposure to the virtual reality simulations, 87 % (20 out of 23) of male participants reported enhanced understanding of molecular biology processes and almost all indicated a move away from memorisation as a learning strategy. In contrast, all but one female participant indicated that they continued to memorise for their subsequent examination. Whilst female participants were engaged in and motivated by virtual reality simulations, this finding clearly signaled that there was something about their learning preferences that these lessons were not catering to. Understanding how to apply virtual reality technology in gender differentiated classrooms is key to stopping female learners from falling into unproductive cycles of non-understanding and memorisation.

3.2 3D Virtual Reality Simulations Neglect Female Learning Preferences

To better understand learning preferences and differences between genders, students' learning behaviours were categorised using both the VARK framework (Fleming and Mills 1992) and the Index of Learning Styles (Felder and Silverman 1988).

In VARK (Fleming and Mills 1992), learners are identified by whether they have a preference for visual learning (V: pictures, movies, diagrams), auditory learning (A: music, discussion, lectures), reading and writing (R: making lists, reading textbooks, taking notes), or kinesthetic learning (K: movement,

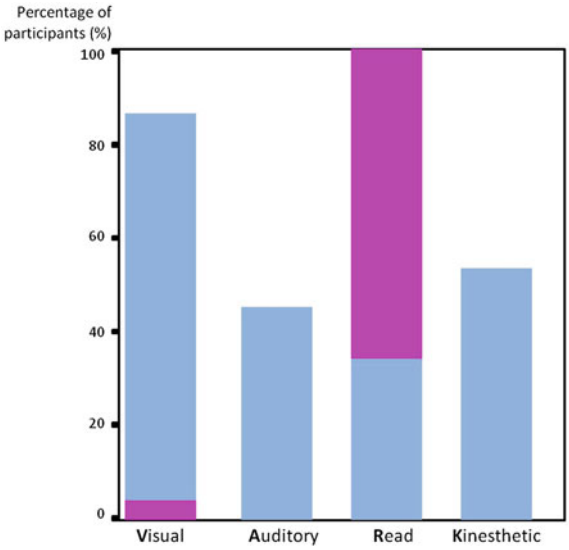


Fig. 1 VARK (Visual, Auditory, Reading and Kinesthetic) learning modalities of female (*coded in pink*) and male participants (*coded in blue*). A large majority of the male participants preferred multiple modalities whilst female participants favoured reading over all other modalities

experiments, hands-on activities). Interestingly, all 23 female participants interviewed exhibited an overwhelming propensity to depend on reading notes, writing and, therefore, memorisation. In contrast, 91 % (21 out of 23) male participants preferred multiple modalities (Fig. 1). These data are similar to previous work gender-based learning styles difference amongst first-year physiology students (Wehrwein et al. 2007), suggesting consistency in the learning profiles of students in biology-related disciplines.

In our current study, the lesson plans for virtual reality-based simulations involve visual simulations, discussions and model constructions (Tan and Waugh 2013). Such activities clearly cater to the multimodality preferences of male students. Female participants might find the activities engaging and novel in comparison to their usual learning routines, but these would not help in information processing and retention for most females. A modified approach for female students might include subtitles to support the animations and narration, as well as notes and incorporation of "quiet time" for reading after viewing the virtual reality simulations. Felder and Silverman (1988) proposed four dimensions of student learning styles, each of which relates to students' preferred modes for receiving information, including (1) the type of information they receive (sensory or intuitive), (2) the modality in which they receive it (visual or verbal), (3) the processes by which they acquire the information (actively or reflectively) and (4) the order in which they receive it (sequentially or globally). These dimensions are useful in considering the diversity of learning styles and how teaching strategies in science classrooms do or do not regularly provide access to learning for these different types of students.

The male participants interviewed were generally sensory, visual and verbal, active and global, while female participants were intuitive, neither visual nor

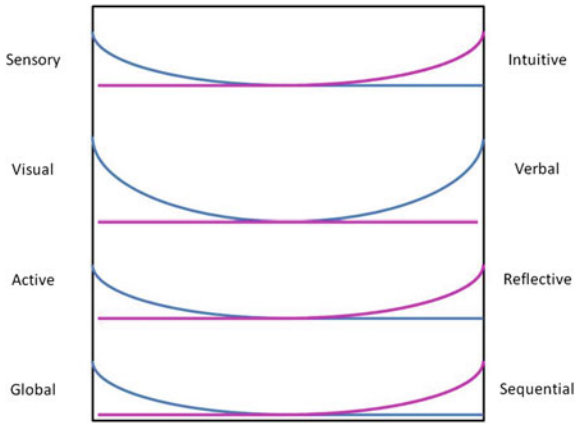


Fig. 2 Learning styles of female (*coded in pink*) and male participants (*coded in blue*). Males and females demonstrated different preferred modes of receiving information (sensory vs. intuitive), the modality in which they received it (visual vs. verbal), the processes through which they acquire new knowledge (active vs. reflective) and the order in which new information is received (global vs. sequential)

verbal, and reflective and sequential (Fig. 2). These learning styles are in line with the VARK (Fleming and Mills 1992) modalities indicated by the participants. For example, the visual and active learning styles of male participants were in line with their preferences for visual and kinesthetic learning modality preferences. Likewise, the reflective learning style of female participants may reflect their preference for quiet reading, while their disinclination towards visual or verbal learning styles is similar to the near absence of female representation in visual and auditory learning modalities in the VARK (Fleming and Mills 1992) framework discussed earlier.

Likewise, the differences in male and female learning styles help explain the success of the former with virtual reality simulations to learn molecular biology. The visual and verbal learning styles of male participants are well suited, therefore, to the animations and narrations used to convey information about molecular biology. It is no wonder that more than 80 % of the male participants indicated that the visuals and narrative used in the study were helpful for understanding. The global learning style of male participants is also in line with the mode of presentation of information. One of the original objectives of three-dimensional virtual reality simulations is to show students how interconnected processes and sub-cellular structures were; the non sequential nature of the information may cause more confusion for female students. This focus on large-scale perspective may also protect male students from the tendency to memorise; a student predisposed to seeing the forest will be hard pressed to recall and regurgitate the names of individual trees.

As with the recommendations generated by VARK (Fleming and Mills 1992) analysis, female students learn better when supported by a slower pace of interactive technology use in the classroom, interspersed with protected time for reading and reflection. Additionally, the sequential learning style of females suggests that animations selected for teaching should present information in logical order. Overall, however, the focus on logical, orderly information, reading and

reflection means that females are particularly prone to the memorisation of details as a learning behaviour. Special effort in the classroom must then focus on helping female students see the "big picture", and to make connections across different learning areas.

4 Conclusions

The present study investigates gender-based differences in learning behaviours, in hopes of explaining differential efficacies of three-dimensional virtual reality simulations in the teaching and learning of molecular biology. While both male and female students demonstrated increased motivation in the subject, only males exhibited a significant improvement in achievement that was directly attributable to the intervention.

These results suggested that males and females may learn differently. Indeed, analysis of learning modalities via the VARK framework (Fleming and Mills 1992) indicated that males were more likely to be multimodal, hence benefiting from the variety of activities linked with lesson plans that used immersive virtual reality for teaching and learning. In contrast, females in the present study preferred to engage only in reading and note-taking, activities not provided for in a virtual reality classroom. Similarly, the learning styles of males and females seem to be in direct opposition on three out of four dimensions of learning styles (Felder and Silverman 1988), with male learning styles more in line with learning using virtual reality technologies.

The results indicate that, while immersive virtual reality simulations are effective for teaching molecular biology, female students should be provided with (1) time and space for reading, reflection and consolidation, (2) skills to help them structure new information logically and sequentially and (3) resources to help them make connections across different areas of the curriculum and focus on this larger perspective of general trends and themes. Plans are underway to test modified lesson plans for female students in technology-intensive biology classrooms in Singapore. It is hoped that these results would contribute to greater understanding of male and female learning behaviours, particularly in technology-intensive learning environments.

Acknowledgments The author would like to thank the National Research Foundation, Singapore, and the Ministry of Education, Singapore, for supporting this study.

This work would not have been possible without the support of many friends and colleagues at Hwa Chong Institution, and Nanyang Girls High School, Singapore. Special thanks also go to Professor Y. Y. Cai from NTU, and Mr. Ngo Boon Keong from Zepth Pte Ltd.

References

Andrich DA (1988) Rasch models for measurement. Sage university paper on quantitative applications in the social sciences, series number 07/068. Sage Publications, Newbury Park

Andrich D, Sheridan B, Luo G (2010) RUMM 2030. RUMM Laboratory, Perth

Bancroft WJ (1995) The two-sided mind: teaching and suggestopedia. Eric DocumentReproduction Service No. ED,384244

Butler D (1994) From learning strategies to strategic learning: promoting self-regulated learning style by post secondary students with learning disabilities. Canadian J Spec Educ 9(3–4):69–101

Cooper SE, Miller JA (1991) MBTI learning-teaching styles discongruities. Educ Psychol Measur 51:699–707

Cronbach LJ, Snow RE (1969) Individual differences in learning ability as a function of instructional variables. Stanford University, School of Education (ERIC 029 001), Stanford

Dwyer KK (1998) Communication apprehension and learning style preference: correlations and implications for teaching. Commun Educ 47(2):137–150

Felder RM, Silverman LK (1988) Learning and teaching styles in engineering education. Eng Educ 78(7):674–681

Fleming ND, Mills C (1992) Not another inventory, rather a catalyst for reflection. Improve Acad 11:137–155

Lundeberg MA, Fox WP, Punochar J (1994) Highly confident but wrong: gender differences and similarities in confidence judgements. J Educ Psychol 8(1):114–121

Mann J (1994) Bridging gender gap: how girls learn. National Association of Elementary School Principals (ERIC 376 611), Arlington

Messick S (1984) The nature of cognitive styles: problems and promise in educational practice. Educ Psychol 19:59–74

Tan S , Waugh R (2013) Use of virtual reality in teaching and learning molecular biology. In: Cai YY (ed) 3D immersive learning in Singapore. Springer, New York

Wehrwein EA, Lujan HL, DiCarlo SE (2007) Gender differences in learning style preferences among undergraduate physiology students. Adv Physiol Educ 31(2):153–157

Transfer of Training of An Educational Serious Game: The Effectiveness of the CASHIER TRAINER

E. A. P. B (Esther) Oprins and J. E (Hans) Korteling

Abstract We investigated the transfer of training of a stand-alone educational serious game supported with automated feedback and instruction compared to conventional on-the-job training (OJT). If transfer of training is sufficiently high, these types of games could reduce time-consuming and expensive OJT saving instructional personnel. In a case study with such a game, the CASHIER TRAINER supported with an ITS, we compared performance and competence of new cashier employees at the workplace in two matched groups: (a) experimental group: acquiring cashier skills in the game CASHIER TRAINER; (b) control group: acquiring cashier skills in conventional OJT. Performance and competence were measured by observation and by self-assessment. The results showed that performance and competence were at least as high and even higher on certain aspects for the CASHIER TRAINER than for the OJT condition. Thus, transfer of training with the CASHIER TRAINER was positive in comparison with OJT. Specifically, non-regular tasks which could be trained with gaming more often and structured than in OJT were performed better. It seems that automated feedback in stand-alone educational serious games for procedural tasks, if well-designed, could replace human tutoring to a certain extent.

Keywords Evaluation · Educational games · Instructional games · Serious games · Intelligent tutoring · Training effectiveness · Instructional effectiveness · Training simulation · Transfer of trainingTransfer of training

E. A. P. B (Esther) Oprins (✉) · J. E (Hans) Korteling
TNO, Delft, The Netherlands
e-mail: esther.oprins@tno.nl

J. E (Hans) Korteling
e-mail: hans.korteling@tno.nl

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0_15,

1 Introduction

The education and training community is increasingly accepting serious games as a potentially valuable, efficient, and effective alternative for conventional forms of education, training, or other applications. Many different types of games exist and the literature does not yet show complete consensus on its definition (e.g., Crookall 2010). Most popular definitions agree, however, that serious games are reality based (i.e., simulations), entertaining, interactive, rule-governed, goal-focused, and competitive (Bell et al. 2008; Tobias and Fletcher 2007; Vogel et al. 2006). They serve other goals than entertainment such as personnel recruitment, change of attitudes, research, selection, concept development, education, and training (e.g., Harteveld 2011, 2012). Within the broader category of serious games, this chapter focuses on the effectiveness of *educational serious games*, which are here referred to as interactive representations of (aspects) of reality intended for learning purposes that deliberately include elements of play or entertainment.

The reviews on studies on transfer of training that have been carried out so far consistently show that evidence is fragmented. Well-designed research on learning effects of educational games is scarce while it is necessary for designing games that are more effective for learning (e.g., Akl et al. 2010; Bekebrede et al. 2011; Connolly et al. 2012; Egenfeldt-Nielsen 2006; Fletcher and Tobias 2008; Freitas 2006; Girard et al. 2012; Hays 2005; Ke 2009; Lee 1999; Leemkuil et al. 2000; O'Neil et al. 2005; Randel et al. 1992; Sitzmann 2011; Vogel et al. 2006; Wouters et al. 2009; Young et al. 2012). Evidence on 'what works' is needed, that is, for which target groups, types of tasks and competences are educational games the optimal alternative and what are the conditions and other factors determining their learning effects? This kind of knowledge will help designers to improve their games (Bedwell et al. 2012). Moreover, for schools, companies, and training organizations it is important to get insight into the factors determining educational effectiveness of games that may be applied in their organizations. One of the applications of educational serious games, as discussed in this article, is to replace (parts of) on-the-job training (OJT). If evidence on transfer of training has been proved, this replacement of OJT will be more easily done by organizations.

Educational serious games are especially useful when conventional OJT does not provide suitable opportunities because of risks caused by errors, costs of using the real system, legislation, sustainability, motivation problems of trainees or limited availability of the operational environment (Korteling et al. 2012). Unlike learning at the workplace, the simulated environment offers the opportunity to train and practice tasks when suitable training opportunities are lacking (Farmer et al. 1999; Oprins 2008). This provides the opportunity to be confronted with learning tasks that do not frequently occur in practice and that can be built up in complexity gradually (Van Merrienboer and Kirschner 2007). In OJT an instructor or supervisor usually coaches the learner while performing tasks in the real-life setting (Oprins et al. 2011). It depends on situations that will happen which learning tasks the learners are confronted with. Educational games are supposed to

be an effective alternative for OJT also because they can (partly) replace this staff and/or give them a more coaching or supervisory role. This is especially possible if games are used that are supported with a well-designed intelligent tutoring system (ITS; Polson and Richardson 1988), providing automated feedback on the actions of the learners. The learners can acquire basic knowledge and skills by themselves to a certain extent and have dealt with learning tasks that do not occur frequently in practice before they enter OJT. This saves instructional personnel at the workplace who supervise learners.

This chapter focuses on transfer of training of educational serious games with artificial instructional tutoring and feedback functions in comparison to OJT. We present a case study with a stand-alone game called CASHIER TRAINER supported with an ITS for new cashier personnel to train their skills on the cash desk in a simulated environment before they enter the supermarket as a possible replacement for (parts of) OJT.

1.1 Transfer of Training

Games may require substantial investments, not only in the design and development of the product itself, but also in the implementation and in education and training of personnel. Therefore, an important question concerns the degree to which educational games are really useful and effective for practical purposes such as job training. Educational games are likely to prove to be a good alternative for OJT if, and only if, they are able to achieve similar or higher learning outcomes at similar or lower expenses. This raises the question of what and how much, relatively speaking, is actually transferred from acquired competence in the game or in OJT to the workplace. This issue is typically captured under the notion of *transfer of training*, that is, skills acquired in a specific training environment may transfer to situations in a different (real world) task environment. After definitions provided by Baldwin and Ford (1988) and Gielen (1995) we define transfer of training in the context of educational gaming as the degree to which knowledge, skills and, attitudes that are acquired by playing a game can be used effectively in real-life (operational, professional) situations. With that, transfer results in a reduction of the amount of OJT required to obtain the training goals (Caro 1977). This definition also refers to the notion of competence that we define as the application of the total combination of acquired knowledge, skills and attitudes effectively in the job (Oprins 2008). This type of research can also be referred to as educational validity (Stainton et al. 2010). External educational validity, as opposed to intern educational validity, involves transfer of training by the game or simulation to the real world.

Educational serious games are supposed to be able to have a relatively high transfer of training in comparison with OJT because of their didactical advantages. From the point of learning theory, a surplus value of educational games is that

competences can be acquired in a realistic, attractive, and challenging manner (Bedwell et al. 2012; Gee 2007; Korteling et al. 2012; Shaffer 2006; Squire 2003). Learners learn actively in authentic (realistic, practical, job-related) and flexible learning environments. Educational games are also assumed to be intrinsically motivating and engaging (Csikszentmihalyi 1990; Malone 1981) and give responsibility to the learner. We suppose that this latter aspect may enhance the learner's self-efficacy (Bandura 1997) and may facilitate the development of self-directed learning (Percival 1996; Stubbé and Theunissen 2008).

These potential didactical and motivational advantages of educational games are in line with modern general learning approaches. In the traditional instructor-centered situation the instructor is dominant. The instructors stand in the center of the attention whereas the learner more or less passively absorbs (lean back) the information that is provided. Contemporary theories of training and instruction, however, conjecture that learners should participate more actively during classroom lessons in a more 'lean-forward' style. Learners should have an active, central role while the instructors should be supportive rather than directive (Johnston and McCormack 1996; Petraglia 1998). Many contemporary constructivist visions of learning and instruction promote these active forms of learning through experience such as discovery learning (Steffe and Gale 1995), action learning (Smith and O'Neill 2003), and experiential learning (Jiusto and DiBiaso 2006). Although constructivism may take many forms (Petraglia 1998; Philips 1998), an underlying premise is that learning is an active process of sense making in which learners seek to build coherent and organized knowledge. Educational gaming may be brought to fit very well to these constructivist notions since it can be used to train relevant competences in a realistic, attractive, and challenging manner, using authentic (realistic, practical, job-related) and flexible learning environments in an active, self-directed way.

1.2 Measurement of Transfer of Training

Empirical studies to get evidence on transfer of training with educational games based on their motivational and didactical features are, however, scarce. Studies on transfer of training are difficult to perform in practice. The degree to which training results in behavioral change on the job is the gold standard of training. It refers to Kirkpatrick's third level of evaluation concerning the enhancement of actual behavior on-the-job based on the training activities (Kirkpatrick 1998). But many studies fall short in measuring training effects beyond Kirkpatrick level 1 (reactions) and 2 (declarative learning; Cohn et al. 2009). Transfer is difficult to measure because it requires research at the workplace that cannot easily be carried out and controlled (Bedwell et al. 2012; Korteling et al. 2013; Salas et al. 2009; Veldhuis and Theunissen 2009).

The previously mentioned reviews of effectiveness studies for educational games that have been done so far (see for a brief qualitative meta-review Harteveld

2012) have shown a large diversity in outcomes and quality of research. The effectiveness of educational games appears to be moderated by the didactical and technical characteristics of the game design as well as the instructional context in which the game is embedded (Sitzmann 2011). More in specific, this refers to, for example, different target groups, tasks, competences, and domains involved as well as types of games. To get insight into these complexities, it is very important to measure process variables related to the learning process itself instead of only measuring outcome variables (Bedwell et al. 2012; Salas et al. 2003, 2009). Process measures examine the manner in which a task is performed by the trainee, whereas outcome measures focus on how well a trainee accomplishes the overall task. Process measures can be a useful diagnostic tools explaining certain outcomes, thus why it happened, illustrating strengths and weaknesses of training program or game that should be either maintained, improved, or further developed to ensure that training goals are met (Cohn et al. 2009; Fowlkes et al. 1999). For practical reasons especially process measures usually have to be based on opinions, questionnaires, ratings, and (self-) evaluations. Direct quantitative measurements in terms of speed, time, or error are mostly more feasible for outcome variables (Korteling et al. 2013), which was also the case in the present experiment.

The reviews also indicate that some studies even do not measure learning effectiveness at all. Some limit themselves to Kirkpatrick's Level 1 (reactions of trainees and experts (e.g., Stehouwer et al. 2005) and others measure other aspects, such as the fidelity of the simulation (Allan et al. 1986). Sometimes they do not measure learning effectiveness appropriately, for instance, they do not use control groups and/or pretests (Sitzmann 2011). Other studies may be hampered by differences between experimental and control groups or in educational contexts of these groups, lack of control groups, test- and selection-effects, short exposure time, weak assessment tests, and preconceived opinions of researchers or teachers (Egenfeldt-Nielsen 2006; Korteling et al. 2013). For these reasons evidence-based studies in which transfer of training has been properly verified are rare (Peck 2012). On the basis of a qualitative meta-review of about 20 review studies on mainly digital serious educational games, Harteveld (2012) concludes that we need to speak of the rise of a potential powerful tool. Gaming has potential, theoretically and based on some of the hints from literature, but we need to figure out how to utilize and proof that potential.

Since transfer of training with games thus has been occasionally proven, getting empirical data on transfer of training, and on the factors that determine this transfer as well, is important to prove the added value of the educational games and to get insight into the most effective didactical, technical, and organizational gaming design properties of specific types of games (Korteling et al. 2013). Generalization of gaming features is restricted; some design features will be more important for particular types of games than others. In this context, the notion *direct transfer* is relevant (Harteveld and Bekebrede 2011). Direct transfer is the more traditional approach with concrete, predefined, and measurable learning objectives to be obtained at the end of the game. In contrast, the open ended learning approach has

more broadly defined, abstract learning objectives and general insights to be achieved. Single-player games are more associated with direct transfer than multiplayer games.

1.3 Instructional Support

One major conclusion of a review of 48 empirical research articles on the effectiveness of educational games (Hays 2005) is that these games seem to be more effective if they are embedded in adequate instructional programs that include debriefing and feedback. In addition, instructional support during play increases the effectiveness of educational games. Teacher guidance and intervention when using games may be carried out by a real teacher or coach facilitating the learning process, in terms of steering the trainee in the right direction and also in providing an effective feedback and debriefing. On the basis of a meta-analytic examination of 65 studies which used a comparison group Sitzmann (2011) draws a similar conclusion concerning the role of instructional support. She deduces that trainees learn more, relative to comparison groups, when so-called simulation games are a supplement to other instructional methods rather than stand-alone instruction. The importance of debriefing while using games is also stressed by others (Alklind Taylor et al. 2012; Crookall 2010; Egenfeldt-Nielsen 2006). A debriefing by human coaches is mostly considered as crucial for learning.

However, when the task to be trained is not very complicated, we suppose that an artificial tutoring system or ITS (Polson and Richardson 1988) may (partly) replace the instructor or coach appropriately. Such automated feedback system may be capable of providing instructions, choosing the right training scenario's, guiding the trainee in the right direction, providing extra information when stuck, and providing feedback and debriefing on task performance. The learning tasks could be built up in complexity in a structured way which helps learning, and also situations that scarcely occur in real-life could be practiced (Van Merrienboer and Kirschner 2007). According to Alklind Taylor (2012), Crookall (2010) and Egenfeldt-Nielsen (2006), human feedback can be partly replaced by automation, but facilitation by training staff which should include a form of debriefing will be indispensable for reflection on what the learners have learned. In line with others (e.g., Farmer et al. 1999; Van Merrienboer and Kirschner 2007) we suppose that the amount of automation in relation to human feedback depends on the type of learning tasks. For instance, in educational games that focus on automating procedures and routines, automated feedback is functional and relatively easy to implement. Automated feedback may be supposed to be less suitable in games intended for learning in complex cognitive tasks (Van Merrienboer and Kirschner 2007). At present debriefing by human coaches is the most effective form of feedback here. Nevertheless, some instruction of human coaches in the form of a debriefing stays important in all types of tasks.

These general insights from the literature are important for all professionals working with educational games or trying to enhance the efficiency of training by saving on instructional personnel. Referring to the previously mentioned conclusions and assumptions, an important issue concerns the beneficial effects of stand-alone educational games provided with intelligent tutoring functions (e.g., Polson and Richardson 1988), which are needed to compensate for the lack of human instructional support and feedback (Egenfeldt-Nielsen 2006). Studying this issue will provide more insight into the potential of stand-alone instructional games and (required) supporting components.

1.4 Goal of The Study

This empirical study aims at getting evidence on transfer of training of particular types of educational serious games. We expect that at least for routine task and procedures to be trained tutoring and feedback can be automated in educational serious games if well-designed. Such a system may sufficiently guide and help students and motivate them to get through the training program. In that case, games can replace (parts of) conventional OJT and this is cost-effective for saving instructional staff at the workplace. Also, the preparation on OJT is efficient since learning tasks could be offered in a well-sequenced way including practicing situations that scarcely occur on the workplace. Following this assumption, we investigated the transfer of training of a stand-alone game for job training with an ITS providing instructions, support cues, and feedback on task performance. We hypothesize that the transfer of training of this type of game could be at least as high as conventional OJT. The game involved in this study is the so-called CASHIER TRAINER for new cashier employees in supermarkets. This game, being representative for educational serious games with an ITS, functions as a case in this study.

2 Cashier Trainer

The CASHIER TRAINER was developed by a leading Dutch training enterprise in the retail sector, Jutten Simulation. This CASHIER TRAINER is being used by various retail organizations in The Netherlands. The game is representative of instructional games with a simulated environment that must be done individually, supported by a virtual tutor. The CASHIER TRAINER is a low-cost virtual training program combining a simulation of a complete task and a simplified 3D representation of the job environment in an attractive and accessible way, see Fig. 1. The game is internet-based and runs on each home computer that is connected to the internet.

Fig. 1 The CASHIER TRAINER (Jutten simulation)

The CASHIER TRAINER is a single-player game that is done independently at home without involvement of a supervisor, coach, or instructor. The primary goal of the learning program is being maximally prepared by practicing the most important learning tasks before the first working day. In addition, the new employees can practice in a safe environment also for tasks they cannot practice very often because they hardly occur in normal situations on the job. The CASHIER TRAINER program is followed by a brief OJT in the supermarket. This provides a check of the employees' skills before starting the real work and gives them the opportunity to discuss subjects learned in the CASHIER TRAINER as a form of debriefing. They get a short introduction followed by coaching by a supervisor while working at the cash desk.

The CASHIER TRAINER is a typical instructional simulation game with an intelligent tutoring system (Polson and Richardson 1988). The CASHIER TRAINER does not only represent the cash desk, as many other related training systems, but it simulates the whole cashier environment including virtual customers. In this way, the game is intended to simulate the entire cashier task and to immerse the

learners in the virtual environment of the supermarket. In the game, new cashier employees follow a strict training program with a well-sequenced order of learning tasks. The CASHIER TRAINER has a sophisticated ITS supporting the learners with explanations, instructions, and feedback on erroneous or incorrect behavior. This feedback is adaptive to the level and type of actions. For instance, the first time that learners execute an incorrect action such as using particular functions of the cash desk, the feedback only consists of 'you are wrong'. After a second mistake, the feedback becomes more concrete and helpful, for instance, 'you uses the wrong button'. A third error follows by very specific feedback, for instance, 'you should use button x'. The errors and feedback are logged and related to a scoring system which is built in as a special game element for motivating and engaging the learners.

The CASHIER TRAINER was developed to train the following learning objectives: registration, operating the cash desk system, scanning of normal and special products, payment procedures (e.g., cash, credit card), safety and control procedures, and communication with clients. These learning objectives are typically procedural so that we assume that feedback could be automated and that the CASHIER TRAINER could replace parts of OJT. The learning tasks are sequenced from part-task to whole task training (Van Merrienboer and Kirschner 2007) and also from only regular tasks to non-regular special tasks. This implies that first only very standard tasks are trained separately, for instance, learners start with scanning regular articles without special assignments such as articles with discount. This is typically part-task training. Next, they combine this with a regular payment method (e.g., cash). Learning tasks become more and more complex, being combined with less regular tasks such as scanning special products (e.g., fruits not weighted) and difficult payment methods (e.g., combination of cash and card). Finally, trainees get mixed training scenarios with all types of regular and non-regular tasks. This can be referred to as whole task training. The level of the trainee determines if he/she has to practice extra on certain tasks or not. Thus, not only the feedback is adaptive but the learning tasks in the game as well.

The quality of this game with regard to simulation fidelity, didactics, game play, and technical aspects were evaluated by gaming and educational experts with the TNO Checklist for Evaluation of Edicational Simulations (TNO-CEES 3.0). This is a new and adapted version of the so-called CONCERT Checklist (Emmerik and Korteling 2002, 2003) that has been developed previously by TNO for the structured evaluation and validation of training simulations. The checklist consists of about 180 items that have been defined on the basis of an extensive literature study on training simulation and Educational gaming (e.g., Korteling et al. 2001). Items include all design, didactical, game play, and fidelity features of the game. Examples of clusters of items are: comprehensiveness of the specification and design process, training program, scenario management, instruction and feedback, intelligent tutoring, game play and mechanics, user interface, models, visual image and content, sound, technical reliance.

The results of this evaluation showed that the CASHIER TRAINER comprised excellent didactical features and human-computer interface features with sufficient physical fidelity and minor technical problems. Major points of possible improvement concerned typical game play features that may be supposed to motivate trainees. The game did not explicitly stimulate or motivate self-directed learning by features that are supposed to enhance entertainment value such as competition, storyline, backstory, or fore shadowing. For the present application of the game for training new cashier employees in retail supermarkets, this extra motivation might not be as important as in other cases, since passing the CASHIER TRAINER program was mandatory for being allowed to start working at the supermarket. In addition, because the cashier task is very procedural as argued, self-directed learning does not seem to be as appropriate here as step-by-step learning with clear corrective feedback provided by the ITS.

In sum, the results of this evaluation were sufficiently convincing that the CASHIER TRAINER could be an appropriate learning device with potentials to replace parts of OJT. In the majority of supermarkets, new cashier employees are only trained by OJT with a supervisor sitting behind the new employee while giving feedback, hints, and instructions. Learning in OJT occurs less structured due to the learning tasks that could not be sequenced and planned as in the simulated environment of the CASHIER TRAINER. This could have consequences for the efficiency of learning in OJT if learners have to wait for non-regular tasks that hardly occur in reality. OJT time could be reduced when the new employees have acquired some basic skills and have experienced non-regular situations already before they enter OJT.

3 Method

3.1 Experimental Design

In order to investigate transfer of training of the CASHIER TRAINER in comparison with conventional OJT, we set up an experiment with two conditions. A total of 45 subjects (7 male, 38 female), who were trained as new cashier employees for a large supermarket chain in The Netherlands, were recruited to participate in the experiment. Subjects were randomly assigned to the control group with conventional OJT ($N = 23$) and the experimental group trained with the CASHIER TRAINER ($N = 22$). Average age was 20 years. An independent t test ($p < 0.05$) did not show significant differences in average age between the two groups.

The OJT group was trained according to the usual training procedures in the retail company. After a short introduction of the general procedures of the cashier task they were trained on the real cash desk under supervision of an experienced cashier employee in the supermarket. The amount of feedback and instruction by

the supervisor depends on the progression of the new employees, on the tasks they have to do, for instance, regular or non-regular groceries, and on the instructional skills of the supervisor. Each subject had another supervisor. The experimental group was trained with only the stand-alone CASHIER TRAINER. This group gets access to the CASHIER TRAINER from home circa 3 weeks before their first working day in the supermarket.

Only for the experiment, we tried to fix the training time for both groups at 3 h in total in order to make the control and experimental groups comparable with each other for research purposes. Then it could be compared how much was learned in the CASHIER TRAINER versus OJT within the same time brackets. However, since the study was executed in the real-life practice of supermarket policy, this training time expressed in minutes turned out to be substantially different for all subjects in the OJT group ($M = 213$; $SD = 129$; $SE = 27$) and in the CASHIER TRAINER group ($M = 188$; $SD = 62$, $SE = 13$). We tried to plan the experiment at a specific time and date to fix OJT duration but this was not always possible for the logistics in the supermarket. In addition, accuracy and policy differs per supermarket; we used many supermarkets in the experiment to get a sufficiently high number of candidates. For the experimental group, the subjects were free in spending time in the CASHIER TRAINER in accordance with the purpose of the game and supermarket policy to put responsibility on the new employees themselves. This explains the high variance between subjects in this group. Unfortunately, we did not have the opportunity to put the experiment separately from the practical operations in the supermarkets.

Another methodological restriction of the experiment was that we were not able to do a pretest but only a post-test. Nevertheless, we could argue that the subjects in both groups were comparable since they did not have any experience as cashier employees at all. They all started at a zero competence level. We controlled for other variables such as education, sex, and age. The post-test was administrated by exactly the same performance test for both groups. This test consisted of handling a customer with a shopping cart of preselected groceries that must be paid. The customer was played by a role-player. The candidates must use the cash desk and apply the correct procedures to handle this customer in the real environment of the supermarket at which the employees were working. In this way an experimentally controlled setting on location was created, measuring transfer from the game to the workplace. In order to be able to make a reliable comparison, the task itself was standardized for all subjects in both control and experimental group. This implies that the same groceries have been put into the shopping cart and the same payments must be done by all subjects. The choice of these tasks was based on a task analysis that was carried out in cooperation with the retail company. The assignment for the candidates included registration on the cash desk, regular articles to be scanned, special articles such as price reduction, showing identity card (ID) for alcohol, vouchers, different payment methods, safety and control procedures.

3.2 Measures

The task analysis resulted in a set of training outcomes that were measured in various ways in order to make the measurements more reliable. Many methods for performance measurement exist (e.g. Oprins 2008; Salas et al. 2009). Typically, reliable and valid measurement of transfer of training to performance at the workplace is not trivial (Egenfeldt-Nielsen 2006; Korteling et al. 2013; Salas et al. 2003). In the present case, the learning outcomes could be measured in a relatively easy and reliable way because the cashier task is a highly procedural task. That is, a task consisting of ordered and discrete action sequences are predefined and well-circumscribed for each situation or event that is encountered. These are executed in a rigid or stereotyped way and lead to a clear predetermined experiment. In this experiment we could measure, rate, and evaluate the handling of shopping carts with payments in a supermarket.

We applied a combination of self-assessment and observations. We made a distinction between *performance* during the test: technical errors directly related to the task, and *competence* in a more generic way: the ability of new cashier employees to do their job. They were measured in multiple ways to enhance reliability. First, two well-trained observers noted the type of errors during the test and counted them, called *observed performance*. This checklist contained 15 types of tasks (items) categorized in five predefined rubrics or scales in accordance with the aforementioned task analysis: registration (2 items, e.g., 'sign up at cash desk'), normal articles (3 items, e.g., 'scanning three cup-a-soup with different taste'), special articles (4 items, e.g., 'bananas without registration code'), payments (4 items, e.g., 'pin card payment'), and control procedures (2 items, e.g., 'asking for ID with alcohol product'). The observers marked each item on the checklist with 0 (wrong) or 1 (correct). The observers also assessed eight general competences at a five points rating scale, called *observed competence*. The ratings ranged from 1 (*poor*) to 2 (*insufficient*), 3 (*moderate*), 4 (*sufficient*), and 5 (*excellent*). The five anchors were further defined in text to be as precise as possible, for instance, anchor 1 (*poor*) was further defined as 'many mistakes, not independent, insecure, slowly'. These observers were educated into the requirements, boundary conditions, and performance criteria of the cashier task. The involvement of two observes makes it possible to calculate inter-rater reliability (Oprins et al. 2006). Finally, the observers measured the time of a particular subtask, that is, scanning 15 regular articles, in order to get insight into the speed of working. This subtask was chosen to make this measurement maximally objective.

Second, the subjects did a self-assessment on the performance variables that are directly related to possible errors during the test, called *self-assessed performance*. Because they are not able to use a checklist while they are doing the test like the observers, they rated themselves on comparable tasks as the observers but they were operationalized in a different way. There were five rubrics: cashier desk (3 items; e.g., 'use of the functionalities of the cash desk'), corrections (3 items; e.g., 'dealing with problems with the cash desk'), scanning (5 items; e.g., 'scanning

articles with bar code'), payments (6 items; e.g., 'counting the money when giving back to the customer'), and control procedures (3 items; e.g., 'checking the bags of customers on stolen groceries'). They were rated on a rating scale, the same five points rating scale with similar anchors as mentioned above. In addition, they rated the same set of 8 competences as rated by the observers, also on the same anchored five points rating scale, called *self-assessed competence*.

In sum, four outcome measures were used as dependent variables in the experiment: observed performance, observed competence, self-assessed performance, and self-assessed competence. We compared the differences between the experimental condition (CASHIER TRAINER) and the control condition (OJT) for these four outcome measures with a Mann–Whitney *U* test (Mann and Whitney 1947). We used this nonparametric test because the majority of the data were not normally distributed as being checked with the Shapiro–Wilk test. We also correlated the four outcome measures with each other to investigate how they relate to each other (Spearman's rho, two-tailed). For all composite variables the medians of the underlying aforementioned scales have been calculated.

The subjects in the CASHIER TRAINER condition also filled out a questionnaire with items rated on the same five points rating scale concerning the CASHIER TRAINER. This questionnaire includes four rubrics or scales: training technology (3 items, e.g., 'I find the CASHIER TRAINER an appropriate learning device for new cashier employees'), the quality of practice (3 items, e.g., 'I find the amount of exercises sufficiently to do the test'), didactics (5 items, e.g., 'I like the corrective feedback if I do something wrong'), and reality (4 items, e.g., 'I feel immersed in the environment of the supermarket'). Additionally, the subjects could give remarks on positive and negative aspects of the game. We investigated possible influences on the outcome variables by calculating correlations (Spearman's rho, two-tailed).

Finally, structured interviews with retail managers, who had worked with both groups of candidates, were carried out. In these interviews we inventoried subjective observations about task performance and types of errors. We also asked for expected savings in training time with the CASHIER TRAINER, both for new cashier employees and for the supervisors needed in OJT although this was not the main question in this study. These data were qualitatively analyzed and only used to explain the quantitative findings in this study.

4 Results

4.1 Observer Assessments

The observers assessed both task performances during the test and general competence. They also measured the duration of the subtask: scanning 15 regular articles. The average speed hardly varied for the OJT group (34.7 s) and CASHIER TRAINER group (35.4 s). Figure 2 presents the median scores of the two groups for the number and type of errors (averaged per subtask; usually only one

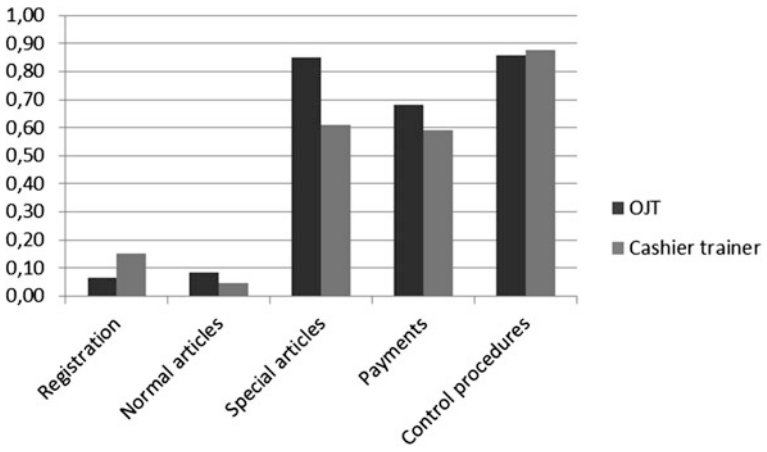


Fig. 2 Median number of errors in observed performance

error could be made per task) made during the performance test, the so-called *observed performance*.

The scores in Fig. 2 are the combined error-scores of the two observers. The reliability was assumed to be sufficiently high: the correlation (Spearman's rho, two-tailed) between the total score of all performance items was .86 significant at $p < 0.001$. The Mann–Whitney U test ($p < 0.05$) did not show significant differences between the two groups. In Fig. 2, we see that the number of errors was lower for regular tasks (registration, normal articles) in comparison with tasks that occur less frequently and are more difficult (special articles, control procedures, payments).

Figure 3 presents the observed median scores on the five-point scale for the two groups on general competence, called *observed competence*.

These are also the combined scores of the two observers. The reliability was assumed to be sufficiently high here as well: the correlation (Spearman's rho, two-tailed) between the total competence scores was 0.69 significant at $p < 0.001$. Figure 3 shows that the scores of candidates in the CASHIER TRAINER condition were slightly superior in 7 out of 8 competences while the scores of the candidates in the OJT conditions were superior only in accuracy. The Mann–Whitney U test ($p < 0.05$) showed that the subtask operating cash desk scored significantly higher for the CASHIER TRAINER than for OJT ($U = 169{,}5$, $p = 0.045$, $r = 0.30$). The rest of them were not significant.

4.2 Self-Assessments

Figure 4 presents the self-assessments of task performance (median scores) of the two groups, OJT and CASHIER TRAINER, called *self-assessed performance*. The variables are not similar to the variables for the observers (see Fig. 2) due to

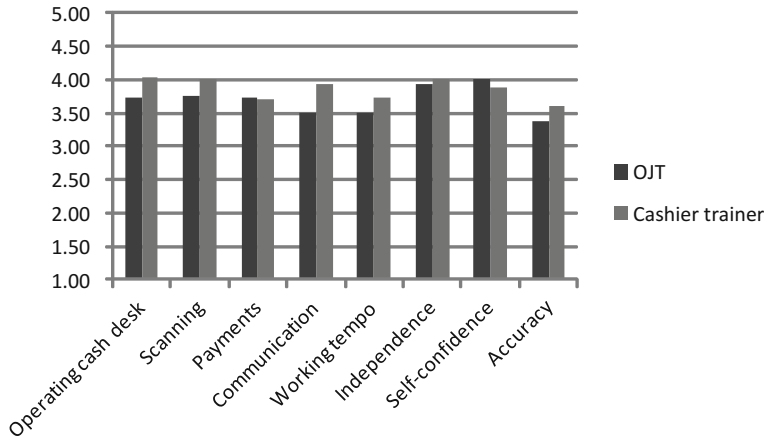


Fig. 3 Median scores for observed competence

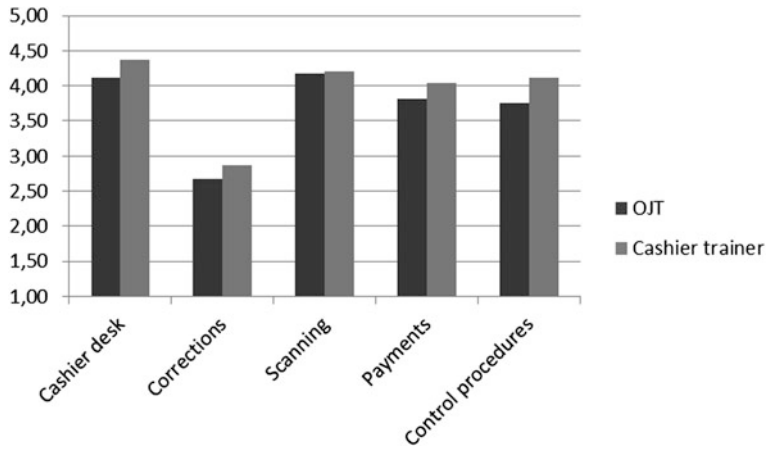


Fig. 4 Median scores for self-assessed performance

the measurement method: the observers counted the errors with value 0 (*wrong*) *or 1* (*correct)* while the subjects rated their task performance on a five points rating scale as stated in the previous section. Figure 4 shows that the candidates in the CASHIER TRAINER condition assessed themselves higher at all five variables than the candidates in the OJT condition. The Mann–Whitney U test ($p < 0.05$) showed that two of the five variables were significantly higher rated in the CASHIER TRAINER group than in the OJT group: payments ($U = 164.0$, $p = 0.041$, $r = 0.31$), and control procedures ($U = 167.0$, $p = 0.047$, $r = 0.30$). The values of the other three variables were not significant. Figure 4 also shows that the candidates rated corrections relatively low in comparison with the other variables.

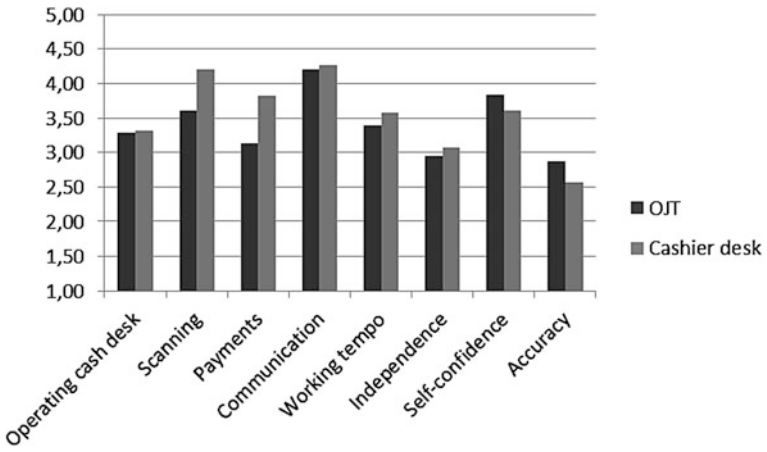


Fig. 5 Median scores for self-assessed competence

Figure 5 presents the self-assessments of general competence (median scores) of both training groups, called *self-assessed competence*. The set of rated competences is similar to the *observed competence* (see Fig. 3) to keep the results comparable. Figure 4 shows that the candidates rated themselves higher at 6 out of 8 competences for the CASHIER TRAINER condition than for the OJT condition. There were two significant differences found with the Mann–Whitney U test ($p < 0.05$): scanning ($U = 152.0$, $p = 0.010$, $r = 0.38$) and payments ($U = 153.5$, $p = 0.018$, $r = 0.35$). It should be noted that the differences between the rated competences are higher for the self-assessments (see Fig. 5) than for the observers (see Fig. 3).

4.3 Total Outcome Measures

Figure 6 presents the median scores of three of the four outcome measures observed competence, self-assessed performance and self-assessed competence. The outcome measures themselves are also based on the medians of the underlying scales. Figure 6 shows that the CASHIER TRAINER group scores slightly higher than the OJT group on the three outcome measures. No significant differences were found with the Mann–Whitney U test ($p < 0.05$). Outcome measure observed performance is not presented in this figure because it is not measured on the same five points rating scale. We did not find any differences between the CASHIER TRAINER group (Mdn = 0.50, SD = 0.048) and the OJT group (Mdn = 0.50, SD = 0.046).

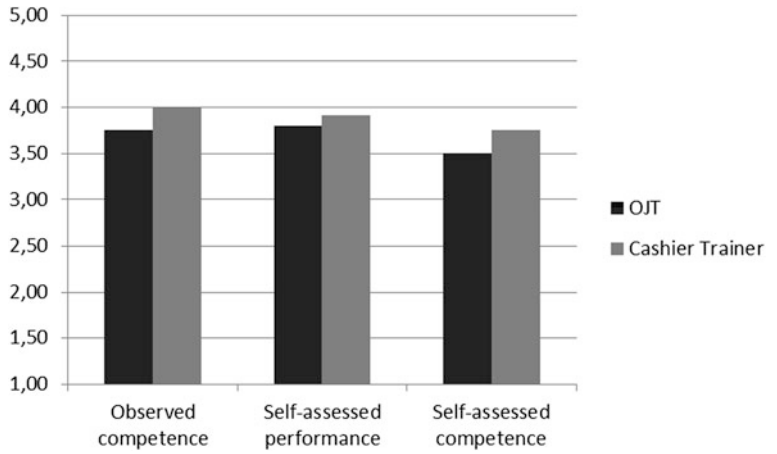


Fig. 6 Median scores for three total outcome measures

4.4 Intercorrelations

Table 1 presents the intercorrelations (Spearman's rho, two-tailed) between the total median scores of the four types of measurements (N = 45):

Table 1 shows that the measurements of performance and competence for the self-assessments are very highly correlated. This implies that performance and competence was strongly related and that the candidates are consistent in their judgments as expected. In the same way, the measurements of performance and competence for the observers are significantly correlated. These correlations are negative because observed performance implies the number of errors which is a reverse scale in comparison to the rating scale from poor to excellent for the other three measures. Observed performance also correlates with the other three measures although not significantly. Interesting is the relatively high and significant correlation between observed and self-assessed competence. This implies that the candidates and the observers have a quite similar picture about their competence. This is less clear for the relationship between observed and self-assessed

Table 1 Correlations (Spearman Rho, 2-tailed) total performance and competence

	1	2	3	4
1. Observed performance	1	−0.68[a]	−15	−0.14
2. Observed competence		1	0.11	0.13
3. Self-assessed performance			1	0.65**
4. Self-assessed competence				1

Note ** $p < 0.01$

Table 2 Correlations (Spearman Rho, 2-tailed) with opinions

	Training technology	Quality of practice	Didactics	Realism
1. Observed performance	−0.03	−0.43*	−0.30	−0.27
2. Observed competence	0.03	0.37	0.11	0.22
3. Self-assessed performance	0.26	0.38	0.20	0.36
4. Self-assessed competence	0.23	0.43*	0.14	0.35

*Note** $p < 0.05$

performance probably due to the different type of measurement, counting errors, or a rating scale.

The candidates' performance was controlled for the possible factors age, education level, gender, and type of job (fulltime, part-time), but no significant correlations were found for any of the four types of measurements, performance or competence. In addition, the amount of time practiced in OJT or in the CASHIER TRAINER was investigated but we did not find significant results for this time factor either.

4.5 Didactical and Motivational Features of the Cashier Trainer

The candidates in the CASHIER TRAINER condition (N = 22) generally had a quite positive opinion on the CASHIER TRAINER. This particularly appears from qualitative data obtained from extra remarks of the subjects. The average ratings of the four scales, rated at the five points scale, varied from 3.7 to 3.9. Table 2 shows the correlations with the various types of measurements. Table 2 shows two significant correlations for quality of practice. In general, the highest correlations although not significant at $p < 0.05$ are found for quality of practice and realism. This suggests that these game features had a positive effect on the outcome of training with the CASHIER TRAINER.

5 Discussion

5.1 Transfer of Training

The objective of the present study was to investigate the transfer of training of the stand-alone CASHIER TRAINER to the workplace in comparison to conventional OJT. The CASHIER TRAINER simulates the entire procedural cashier task, immerses the learners in the virtual environment of a supermarket, and is fitted out with an ITS. The CASHIER TRAINER is therefore representative for single-player educational games with such an automated instruction and feedback

functionality. Training outcomes of the experimental group trained with the CASHIER TRAINER were compared to a control group trained conventionally on-the-job. The two groups executed the same performance test at the workplace. Both observed and self-assessed performance as well as observed and self-assessed competence were measured. Transfer of training in this study refers to the (positive) training outcomes of the CASHIER TRAINER group relative to the OJT group. The results show that the outcomes of CASHIER TRAINER candidates are at least as high as the outcomes of the OJT candidates (see Figs. 2, 3, 4, 5 and 6). Although the presented figures suggest that the new employees trained with the CASHIER TRAINER score higher on the most outcome variables than the OJT group, the results of only some of the variables were significant. The overall outcome measures are generally a little bit higher for the CASHIER TRAINER group than for the OJT group but these small differences were not significant either (see Fig. 6).

From these results we may conclude that the new employees in the CASHIER TRAINER condition have learned at least as much as the OJT candidates and even more on certain aspects. We may assume that its automated feedback and instructional system has done its work appropriately; the candidates did not have a human coach when playing the CASHIER TRAINER in contrast with OJT. Thus, transfer of training with the CASHIER TRAINER to the workplace is definitely positive. As argued, we tried to fix the training time in the two conditions to be really able to compare them with each other but we saw that the variance between subjects was very high. Nevertheless, the OJT time was on average somewhat higher than in the CASHIER TRAINER. We did not found any effects of training time on the subjects' individual performance or competence. Therefore, we could derive that the experimental and control group generally have learned as much and that this was probably not caused by a shorter training time.

The practical implication of these findings is that we can assume that the CASHIER TRAINER can replace a certain part of total OJT time. Due to its artificial feedback system, candidates can do this game at home, independently of a human coach, before they enter the supermarket. Therefore, the CASHIER TRAINER can save on teacher or coaching capacity needed in conventional OJT. In the interviews, the supermarket managers estimated these savings on instructional personnel at about 50–70 % but this estimation has not been quantified yet. This does not mean that we can assume that the OJT could be fully replaced by the CASHIER TRAINER; this did not belong to the scope of this study. We did not investigate how much OJT time could be saved exactly.

Considering the cashier trainer as representative educational games with an ITS, this finding about the expected transfer of training of these types of educational games and their possible replacement of OJT time agrees with the expectations found in the literature. Educational games may show high training effectiveness if they are supported with appropriate instructional support (e.g., Egenfeldt-Nielsen 2006; Farmer et al. 1999; Hays 2005; Sitzmann 2011). In this study, we focused on games with an automated feedback system and we did not explicitly investigate the (extra) role of human coaches, for instance, in the form of

a debriefing as a follow-up of the game (Alklind Taylor et al. 2012; Crookall 2010). It would be interesting to compare automated and human instruction, feedback, and debriefings with each other in another study to get a deeper insight into the possible surplus value of automation by gaming versus OJT by human supervisors.

The present study was a typical case study with an experimental design for measuring transfer of training with an experimental and control group and executed in real-life working environment (Sitzmann et al. 2011; Stainton et al. 2010). This has caused some methodological restrictions. One of them involves problems with fixing the training time as explained. In addition, due to practical obstacles, we had to leave out a pretest. This would make it possible to compare the delta of learning. From a methodological point of view this is certainly preferred. Moreover, the number of subjects involved was relatively low which makes it even harder to find clear (significant) results. Next, the performance test only consisted of one case with one customer due to practical time restrictions. Two tests would be more reliable. This study should therefore be repeated in a pre and post-test experimental design with fixed training conditions including time, multiple performance tests, and a higher amount of subjects to get more valid evidence.

5.2 Type of Tasks

To get more insight into the reasons why training effectiveness appeared to be higher in the CASHIER TRAINER condition, we looked for some trends in the performance and competence data presented in Figs. 2, 3, 4 and 5. The number of errors on relatively easy and regular tasks, such as scanning normal articles and registration at the cash desk, were relatively low (see Fig. 2). The significant differences between experimental and control group that we found, especially for self-assessment (see Figs. 4 and 5) could be referred to as less regular tasks. This is particularly true for payments. In the performance test as well as in the real-life world, these payments are rather complicated for new cashier employees when not standard payment methods are used, for instance, a combination with pin and cash as one of the tasks in the tests. Also back-calculating money is difficult for the most new employees according to the supermarket managers. The CASHIER TRAINER offers the opportunity to practice with a variety of payment methods, from simple to complex, in a safe training environment, including payments that only scarcely occur in reality. We did also find significant results for self-assessment on scanning. This measure comprises also non-regular tasks such as scanning special articles (e.g., price reduction, alcohol). The third significant effect was found for self-assessment on control procedures which again happen less frequently (e.g., asking for ID). Scanning non-regular articles and control procedures can also be systematically trained and practiced in the CASHIER TRAINER. This is much more difficult in the real-life environment of OJT.

These findings provide evidence on earlier statements in the literature that educational games offer the possibility to practice these type of non-regular tasks more often in a well-structured sequence of dedicated scenarios: one of the reasons to use gaming or simulation instead of training in real-life environments (Farmer et al. 1999; Korteling et al. 2012; Oprins 2008; Oprins et al. 2011; Van Merrienboer and Kirschner 2007). The ITS (Polson and Richardson 1988) of the CASHIER TRAINER offers systematic instruction and feedback to learn these non-regular and also relatively difficult tasks systematically and effective in a safe and controlled environment with automated instruction and feedback (Korteling et al. 2012; Oprins 2008).

In our study the learning objectives were predefined and relatively easy to measure, referring to as direct transfer (Harteveld and Bekebrede 2011). In contrast to many other professional tasks, this was possible because the task of cash desk employees is strongly procedural (e.g., Van Merrienboer and Kirschner 2007). This means that the task is based on discrete action sequences, executed in a rigid or stereotyped way and leading to a clear predetermined goal. The CASHIER TRAINER is a stand-alone device developed to train this highly procedural cashier task. On the basis of the positive results so far, we may conclude that educational games—if well-designed and equipped with a well-designed ITS—are effective for this kind of procedural tasks. If other kinds of tasks and learning objectives are involved such as open ended cognitive tasks or complex perceptual-motor tasks (Farmer et al. 1999; Van Merrienboer and Kirschner 2007), it may be more difficult to obtain high positive transfer from these types of educational games with an ITS. It is the question whether human feedback can be replaced by automation with these types of tasks (Alklind Taylor 2012; Egenfeldt-Nielsen 2006).

5.3 Future Learning at the Workplace

Although transfer of training to the workplace is positive, a certain amount of OJT is supposed to be needed after having finished the CASHIER TRAINER program. Especially, the various self-assessed competences were rated differently by the candidates (see Fig. 5). These findings suggest that the candidates are relatively less confident in operating the cash desk in comparison with scanning and communication with customers and theirs working speed, independence, and accuracy had to be further developed according to themselves. This is an expectable outcome since only 3 hours of training is rather low for building routine in any training (e.g., Van Merrienboer and Kirschner 2007). In addition, the retention time must be taken into account here. The time between finishing the CASHIER TRAINER program and the first working day (at which the performance test took place) was about 2 to 3 weeks. This increased the risk of skill loss during this retention period and thus may lead to lower transfer. This risk counts especially for procedural tasks that must be practiced often in order to maintain routine. In general, procedural tasks must be frequently practiced and repeated just before

entering OJT in order to achieve maximal transfer (Van Merrienboer and Kirschner 2007). If the time gap between the CASHIER TRAINER program and the first working day had been smaller, for instance, only a few days, the differences in transfer of training with the OJT conditions might have been much higher. For the control group, the performance test was immediately following OJT. This was a rather high contrast with the CASHIER TRAINER which unfortunately we could not control for.

Next to routine in task performance, also some more general competences are supposed to be acquired during prolonged working at the workplace. For instance, the stress component, for instance, rows of waiting customers, is usually missing in instructional games or simulations such as the CASHIER TRAINER. Psychosocial skills such as coping with stress are thus more difficult to train with gaming or simulation because they require face to face or life training (Farmer et al. 1999). In addition, specific communication procedures could be trained in the CASHIER TRAINER but the game did not include natural speech. The virtual customers in the CASHIER TRAINER had a limited arsenal of sentences that could be answered by the candidates in a multiple choice menu. Nevertheless, both the observers and the candidates rated communication as rather high both in the CASHIER TRAINER and OJT condition. Apparently, the virtual customers could replace real-life customers for the acquisition of a specific subset of communication procedures despite of a lack of natural speech. This is particularly relevant in domains where practicing with real customers or clients is difficult to realize, for instance, virtual patients are often used within the medical domain (Akl et al. 2010).

5.4 Observation Versus Self-Assessment

We also investigated the interrelations between the four outcome variables. Within candidates and within observers, the correlations between performance and competence were very high as expected since they refer to their own judgments. Both the performance and competence ratings were based on the same task execution. The relationship between self-assessed competence and observed competence seems stronger than between self-assessed performance and observed performance. The main reason is probably that these two measurement methods were different: counting the formal errors with only two possible scores, correct or wrong, versus more intuitive self-assessing performance more generally rated on a five points scale. It supports the idea that multiple measures should be used to get reliable assessments that can be used for performance measurement in simulation based training (Oprins 2008; Salas et al. 2003, 2009).

It is remarkable that the differences between the various self-assessed competences (see Fig. 5) were higher than for the observed competences (see Fig. 3). This might be explained by the fact that the possibility to get insight into own performance of the trainees by introspection was lacking for the observers. For instance, only the

candidates themselves could have a precise feeling about their self-confidence in operating the cash desk independently of the supervisor. The observers could only rely on what they saw. This difference in observation perspective may also have increased the probability of *halo errors*: the tendency to think of a person in general as rather good or inferior and to color the ratings on specific dimensions by this general feeling (Oprins et al. 2006; Oprins 2008; Thorndike 1920). Probably, the observers' ratings were influenced by this well-known rating error.

5.5 Didactical and Motivational Features

Finally, the candidates' opinion on the didactical and motivational features of the CASHIER TRAINER was related to observed and self-assessed performance and competence. In general, the candidates were quite positive on these features. This appears also from the qualitative remarks on the questionnaire as well as from the interviews. The new cashier employees were motivated to do the CASHIER TRAINER, they liked the tutoring and feedback system, and they felt immersed in the simulated environment of the supermarket. These results are consistent with the notion that games are motivating (Csikszentmihali 1990; Malone 1981). They also agree with the idea that fidelity in particular is a very important feature of simulation (Allan et al. 1986; Farmer et al. 1999; Korteling et al. 2012). We would expect that comparable results may be found for other games although this study is restricted to only one case.

However, only the candidates' general opinion on the CASHIER TRAINER was used without really measuring the underlying learning processes such as self-directed learning, way of practicing, repetition of exercises, etcetera. Measuring process variables in combination with outcome variables is crucial to get insight into the factors determining transfer of training: what works and what doesn't work (e.g., Bedwell et al. 2012; Salas et al. 2009) for which groups, tasks, and conditions. Therefore, the relationship between the fidelity, didactical and motivational features and learning processes should be investigated in more detail in the future. In order to get more insight into the most effective design characteristics of instructional simulation games, this kind of research should be based on a larger set of comparable educational games. This conclusion confirms similar kinds of suggestions found in previous reviews (e.g., Ack et al. 2010; Ke 2009; Lee 1999; Randel et al. 1992; Sitzmann 2011; Vogel et al. 2006).

5.6 General conclusion

This is one of the first empirical studies in which real evidence was collected for transfer of training for a stand-alone educational serious game equipped with a well-designed ITS compared to conventional OJT. We found that the amount of

learning was at least as high for the candidates who did the CASHIER TRAINER as for candidates who did OJT. Since transfer of training of this representative educational game was sufficiently high; we can conclude that these types of procedural stand-alone games, if well-designed, could reduce the amount of relatively expensive OJT. For particular procedural tasks, an appropriate ITS can replace teachers, coaches or human supervisors, necessary to guide the trainee through the training program and to provide feedback. If this reduces OJT time, this may result in substantial savings on costs of instructional staff. We suppose that this finding may be promising for future development of educational games for similar types of procedural tasks. A simulated environment offers the opportunity to practice both regular as well as non-regular tasks in well-sequenced immersive scenarios. This is not possible in OJT where the learning tasks cannot always easily be planned in advance if non-regular tasks hardly happen in real-life. This serves as an additional advantage of these types of educational serious games.

Acknowledgments This study was a Knowledge Transfer Project of the research program Games for Training and Entertainment (GATE) which was supported by the Dutch ICT-Regie and NWO. We would like to thank Jutten Simulation for using the game CASHIER TRAINER as a case in this study. We also would like to thank the retail company Sligro Foodgroup n.v. in The Netherlands for making it possible to collect data in their EMTÉ supermarkets.

References

Akl EA, Pretorius RW, Sackett K, Scott Erdley W, Bhoopath P, Alfarah Z, Nemann HJ (2010) The effect of educational games on medical students' learning outcomes: a systematic review. BEME guide no 14

Allan JA, Hays JT, Buffardi LC (1986) Maintenance Training Simulator Fidelity and Individual Differences in Transfer of Training. Hum Factors 28:297–509

Alklind Taylor AS, Backlund P, Niklasson L (2012) The Coaching Cycle: A Coaching-by-Gaming Approach in Serious Games. Simul Gaming 20(10):1–25

Baldwin TT, Ford JK (1988) Transfer of training: a review and directions for future research. Pers Psychol 41:63–105

Bandura A (1997) Self-efficacy: the exercise of control. Freeman, New York

Bedwell WL, Pavlas D, Heyne K, Lazzara EH, Salas E (2012) Toward a taxonomy linking game attributes to learning: an empirical study. Simul Gaming 43(6):729–760

Bekebrede G, Warmelink HJG, Mayer IS (2011) Reviewing the need for gaming in education to accommodate the net generation. Comput Educ 57(2):1521–1529. doi:10.1016/j.compedu.2011.02.010

Bell BS, Kanar AM, Kozlowski SWJ (2008) Current issues and future directions in simulation-based training in North America. Int J Hum Res Manag 19:1416–1436

Caro, P.W. (1977). Some factors influencing air force simulator training effectiveness HUMRRO Technical Report tr-77–2. Alexandria, Virginia: Human Resources Research Organisation

Cohn J, Kay S, Milham L, Bell Carroll M, Jones D, Sullivan J, Darken R (2009) Training effectiveness evaluation: from theory to practice. In: Schmorrow D, Cohn J, Nicholson D (eds). The PSI handbook of virtual environments for training and education,pp 157–172

Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM (2012) A systematic literature review of empirical evidence on computer games and serious games. Comput Educ 59(2):661–686. doi:10.1016/j.compedu.2012.03.004

Crookall D (2010) Serious games, debriefing, and simulation/gaming as a discipline. Simul Gaming 41(6):898–920
Csikszentmihali M (1990) Flow: the psychology of optimal experience. Harper and Row, New York
Egenfeldt-Nielsen S (2006) Overview of research on the educational use of video games. Digital Kompetanse 1(3):184–213
Emmerik ML, van, Korteling JE (2002) *Certificering van trainingssimulatoren 2*: de TNO-TM checklist. [Certification of training simulators 2: The TNO-TM checklist] Report TM-02-D010. Soesterberg, The Netherlands: TNO Human Factors research Institute
Emmerik ML, van, Korteling JE (2003) Certificering van trainingssimulatoren 3: Computer-gebaseerd ONdersteuningmiddel voor CERtificering van Trainingssimulatoren. [Certification of training simulators 3: Computer-based Support Tool for Certification of Training Simulators] Report TM-03-D005. Soesterberg, The Netherlands: TNO Human Factors research Institute
Farmer E, van Rooij J, Riemersma J, Jorna P, Moraal J (1999) Handbook of simulator-based training. Ashgate, Aldershot
Fletcher JD, Tobias S (2008) What research has to say (thus far) about designing computer games for learning. Paper presented at the American Educational Research Association, New York
Fowlkes J E, Dwyer DJ, Milham L M, Burns J J, Pierce L G (1999) Team skills assessment: A test and evaluation component for emerging weapon systems. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL
Freitas S, de (2006) Learning in immersive worlds: a review of game-based learning (Tech. Rep.). Joint Information Systems Committee Bristol, Bristol
Gee JP (2007) What videogames have to teach us about learning and literacy. Palgrave Macmillan, New York
Gielen EWM (1995) Transfer of training in a corporate setting (doctoral thesis). University Twente, Enschede
Girard C, Ecalle J, Magnan A (2012) Serious games as new educational tools: How effective are they? A meta-analysis of recent studies. J Computer Assist Learn. doi:10.1111/j.1365-2729.2012.00489.x
Harteveld C (2011) Triadic game design: Balancing reality, meaning and play. Springer, London
Harteveld C (2012) Making sense of Virtual Risks: a Quasi-Experimental Investigation into Game-Based Training. IOS Press, Amsterdam
Harteveld C, Bekebrede G (2011) Learning in Single-Versus Multiplayer Games: The More the Merrier? Simul Gaming 42(1):43–63
Hays RT (2005) The effectiveness of instructional games: a literature review and discussion. Technical Report 2005-004. Naval Air Warfare Training Systems Division. Orlando, U.S.A
Jiusto S, DiBiasio D (2006) Experiential learning environments: Do they prepare our students to be self-directed, life-long learners? J Eng Educ 95:195–204
Johnston S, McCormack C (1996) Integrating information technology into university teaching: Identifying the needs and providing the support. Int J Educ Manag 10(5):36–42
Ke F (2009) A qualitative meta-analysis of computer games as learning tools. In Ferdig RE (ed) Handbook of research on effective electronic gaming in education vol I. pp 1–32
Kirkpatrick DI (1998) Evaluating Training Programs: The Four Levels, 2nd edn. Berrett-Koehler, San Francisco
Korteling JE, Helsdingen AS, Theunissen NCM (2012) Serious Games @ Work Learning job-related competences using serious gaming. In: Bakker A, Derks D (eds) The Psychology of Digital Media at Work. London, New York: Psychology Press LTD/Taylor and Francis Group
Korteling JE (2012) Evaluation of a cashier trainer: true evidence for transfer of training of serious gaming. In: Veltkamp R (ed) Growing knowledge for games utrecht NL: Control Magazine 68–69
Korteling JE, Oprins EAPB, Kallen VL (2013) Measurement of Effectiveness for training simulations. In: Wang Z (ed) RTO-MP-SAS-095 Cost-Benefit Analysis of Military Training. Paper presented at the SAS Workshop held in Amsterdam, The Netherlands, 5–6 June 2012. [http://www.cso.nato.int/abstracts.aspx]

Korteling JE, Padmos P, Helsdingen AS, Sluimer RR (2001) *Certificering van trainingssimulatoren 1: kennisinventarisatie* [Certification of training simulators 1: knowledge inventarization] Report TM-01-D003. TNO Human Factors Research Institute, Soesterberg, The Netherlands

Leemkuil H, de Jong T, Ootes S (2000) Review of educational use of games and simulations (IST-1999-13078 Deliverable D1). University of Twente, Enschede

Lee J (1999) Effectiveness of a computer-based instructional simulation: a meta-analysis. Int J Instr Media 26:71–85

Malone TW (1981) Towards a theory of intrinsically motivating instruction. Cognitive Sci 4:333–369

Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 18(1):50–60

van Merrienboer JJG, Kirschner P (2007) Ten steps to complex learning: a systematic approach to four-component instructional design. Lawrence Erlbaum Associates, Mahwah

O'Neil HF, Wainess R, Baker EL (2005) Classification of learning outcomes: evidence from the computer games literature. Curric J 16(4):455–474. doi:10.1080/09585170500384529

Oprins E (2008) Design of a competence-based assessment system for air traffic control training. Maastricht University, Doctoral dissertation

Oprins E, Burggraaff E, Roe R (2011) Analysis of learning curves in the on-the-job training of air traffic controllers. In: D'Oliviera TC (ed) Mechanisms in the chain of safety. Ashgate Publishing Company, Aldershot

Oprins E, Burggraaff E, Van Weerdenburg H (2006) Design of a competence-based assessment system for air traffic control training. Int J Aviat Psychol 16(3):297–320

Oprins E, Burggraaff E, Van Weerdenburg H (2008) Reliability of assessors' ratings in competence-based air traffic control training. Hum Factors Aerosp Saf 6(4):305–322

Peck M (2012) Tools or toys? Training games are popular, but no one knows how well they work. Train simul J dec2011/jan2012

Percival A (1996) Invited reaction: An adult educator responds. Hum Resour Dev Quart 7:131–139

Petraglia J (1998) Reality by design: the rhetoric and technology of authenticity in education. Erlbaum, Mahwah

Phillips DC (1998) How, why, what, when, and where: Perspectives on constructivism in psychology and education. Issues Educ 3:151–194

Polson M, Richardson J (eds) (1988) Foundations of Intelligent Tutoring Systems. Lawrence Erlbaum Associates, Hillsdale

Randal JM, Morris BA, Wetzel CD, Whitehill BV (1992) The effectiveness of games for educational purposes: a review of recent research. Simul Gaming 23:261–276

Roscoe SN, Williges BH (1980) Measurement of transfer of training. In: Roscoe SN (ed) Aviation Psychology. The Iowa State University Press, Iowa

Salas E, Milham LM, Bowers CA (2003) Training evaluation in the military: misconceptions, opportunities, and challenges. Mil Psychol 15:3–16

Salas E, Rosen MA, Held JD, Weissmuller JJ (2009) Performance Measurement in Simulation-Based Training: A Review and Best Practices. Simul Gaming 40(3):328–376

Sitzmann T (2011) A meta-analytic examination of the instructional effectiveness of computer-based simulation games. Pers Psychol 64:489–528

Smith PAC, O'Neil J (2003) A review of action learning literature 1994-2000. Part 1: Bibliography and comments. J Workplace Learn 15:63–69

Stainton AJ, Johnson JE, Borodzicz EP (2010) Educational Validity of Business Gaming Simulation: a research methology framework. Simul Gaming 41(5):705–723

Steffe L, Gale J (eds) (1995) Constructivism in education. Lawrence Erlbaum Associates, Inc., Hillsdale

Stehouwer M, Serné M, Niekel C (2005) A tactical trainer for air defence platoon commanders. In: Proceedings of the interservice/industry, training, simulation, and education conference. Orlando I/ITSEC 2005, Paper no. 206

Shaffer (2006) How computer games help people learn. Palgrave Macmillan, New York
Squire K (2003) Video games in education. Int J Intell Simul Gaming 2(1):49–62
Stubbé HM, Theunissen NCM (2008) Self-directed learning in a ubiquitous learning environment: a meta-review. In: Proceedings of special track on technology support for self-organised learners 2008, pp 5–28
Thorndike EL (1920) A constant error in psychological ratings. J Appl Psychol 4:25–29
Tobias S, Fletcher JD (2007) What research has to say about designing computer games for learning. Educ Technol 47:20–29
Veldhuis GJ, Theunissen NCM (2009) Transfer of Training. Onderzoek naar een maximaal leereffect [Transfer of training. Research on a maximal learning effect]. Opleiding en Train 11:20–22
Vogel J, Vogel DS, Cannon-Bowers J, Bowers CA, Muse K, Wright M (2006) Computer gaming and interactive simulations for learning: a meta-analysis. J Edu Comput Res 34:229–243
Wouters P, van der Spek E, Van Oostendorp H (2009) Current practices in serious game research: a review from a learning outcomes perspective. Games-based learning advancements for multi-sensory human computer interfaces: techniques and effective practices 232–250
Young MF, Slota S, Cutter AB, Yukhymenko M (2012) Our princess is in another castle: A review of trends in serious gaming for education. Rev Educ Res 82(1):61–89

Index

Y. Cai and S. L. Goei (eds.), *Simulations, Serious Games and Their Applications*, Gaming Media and Social Effects, DOI: 10.1007/978-981-4560-32-0,