2nd ASIAN CONFERENCE

ON

INTELLIGENT GAMES AND SIMULATION

GAME-ON'ASIA 2010

2nd ASIAN SIMULATION

TECHNOLOGY CONFERENCE

ASTEC'2010

EDITED BY

Wenji Mao

and

Lode Vermeersch

MARCH 1-3, 2010

Westin Bund Center
Shanghai
CHINA

A Publication of EUROSIS-ETI

2nd Asian Conference
on
Intelligent Games and Simulation

2nd Asian Simulation
Technology Conference

SHANGHAI, CHINA

MARCH 1 - 3, 2010

Organised by

ETI

Sponsored by

EUROSIS

Co-Sponsored by

Ghent University

KBC

Hosted by

Westin Bund Center

Shanghai, China

ASTEC EDITORS

General Conference Chair

Lode Vermeersch, KBC Gold State FMC, Co., Shanghai, China

Past Conference Chairs

Ruck Thawonmas, Ritsumeikan University, Shiga, Japan
Frank Rinaldo, Ritsumeikan University, Shiga, Japan
Kozaburo Hachimura, Ritsumeikan University, Shiga, Japan

INTERNATIONAL PROGRAMME COMMITTEE

Simulation Methodology

Hamid Demmou, LAAS/CNRS, Toulouse Cedex France
Helge Hagenauer, Universitaet Salzburg, Salzburg, Austria
David Hill, Universite Blaise Pascal, Aubiere, France
Vladimir Janousek, Brno University of Technology, Brno, Czech Republic
Xiaolin Hu, Georgia State University, Atlanta, USA
Panajotis T. Katsaros, Aristotle University Thessaloniki, Thessaloniki, Greece
Henrikas Pranevicius, Kaunas University of Technology, Kaunas, Lithuania
Damien Trentesaux, LAMIH/SP, le mont Houy, Valenciennes cedex, France

Simulation Tools

Xiaochen Li, Institute of Automation, Chinese Academy of Sciences, China
Renate Sitte, Griffith University, Gold Coast, Australia
Alfonso Urquía, UNED, Madrid, Spain

Agent Based Simulation

Marina Bagic, University of Zagreb, Zagreb, Croatia
Andre Campos, Natal RN, Brazil
Joel Colloc, Universite Lumiere Lyon 2, Lyon, France
Jacinto Davila, Universidad de Los Andes. Merida. Venezuela
Julie Dugdale, Institut IMAG, Grenoble Cedex, France
Alfredo Garro, Universita' della Calabria, Arcavacata di Rende, (CS)Italy
Wenji Mao, Chinese Academy of Sciences, Beijing, P.R. China
Juan Merelo, Telefonica, Spain
Lars Moench, FernUniversitaet in Hagen, Hagen, Germany
Andreea Monnat, University of Luxembourg, Luxembourg
Michael J. North, Argonne National Laboratory, Argonne, USA
Eugenio Oliveira, Faculdade de Engenharia da Universidade do Porto, Porto, Portugal
Isabel Praca, Instituto Superior de Engenharia do Porto, Porto, Portugal
Maria Joao Viamonte, Instituto Superior de Engenharia do Porto, Porto, Portugal
Bejrouz Zarei, Management College of Tehran University, Tehran, Iran

Aerospace Simulation

Nima Amanifard, University of Guilan, Rasht, Iran
Ernst Kesseler, NLR, Amsterdam, The Netherlands
Zdravko Terze, University of Zagreb, Zagreb, Croatia
Kai Virtanen, Helsinki University of Technology, HUT, Finland

Engineering Simulation

Chrissanthi Angeli, Technological Education Institute of Piraeus, Athens, Greece
Jacques Andre Astolfi, Brest Armess, France
Josko Dvornik, University of Split, Croatia
Rene Heinzl, TU Wien, Vienna, Austria
Feliz Teixeira, University of Porto, Porto, Portugal
Morched Zeghal, National Research Council Canada, Ottawa, Canada

INTERNATIONAL PROGRAMME COMMITTEE

GAME ON'ASIA
2010

ASTEC'2010

All articles published in these Proceedings have been peer reviewed

EUROSIS-ETI Publications are Thomson-Reuters and INSPEC referenced

For permission to publish a complete paper write EUROSIS, c/o Philippe Geril, ETI Executive Director, Greenbridge NV, Wetenschapspark 1, Plassendale 1, B-8400 Ostend, Belgium.

EUROSIS is a Division of ETI Bvba, The European Technology Institute, Torhoutsesteenweg 162, Box 4, B-8400 Ostend, Belgium

Printed in Belgium by Reproduct NV, Ghent, Belgium
Cover Design by Grafisch Bedrijf Lammaing, Ostend, Belgium

EUROSIS-ETI Publication

ISBN: 978-9077381-54-0
EAN: 978-9077381-54-0

Preface

Welcome to the second joint Game-On-Asia, our future annual Asian Conference on Simulation and AI in Games, and ASTEC, our second future annual Asian Simulation Technology Conference. On behalf of all the people who made this joint conference happen, we wish to welcome you to this special Asian event.

During the past years, Game-On has offered in other parts of the world, namely Europe and North-America, an opportunity for researchers and practitioners to present their findings and research results in the new and exciting field of gaming technologies, while on the other hand, the ASTEC conference aims to become the counterpoint of the ISC Conference in Europe and NASTEC Conference in North-America. This year, the joint event returns to Asia after a three years absence with a technical program covering on the gaming side, an overview of new technology, gamer behaviour and classification issues and game design, while on the ASTEC side focusing on Complex Systems Simulation and Transport Simulation.

All the contributed papers to both conferences have undergone a serious paper review and helped us to achieve our goal of creating a well-balanced informative and topical conference. Special recognition goes to each of the contributing authors for their dedication and effort in their field of research. In addition to all the accepted papers, we assembled a program comprising two keynote speeches (given by Lode Vermeersch and Prof. Fei-Yue Wang), and two invited speeches (given by Zhiyuan Luo and Mei Si).

On behalf of the Organizing Committee, we would like to extend our personal thanks to all the members of the International Program Committee for their hard work in reviewing and selecting the best papers to be presented from all the received submissions. The success of this conference is credited to them, as well as to the session chairs, presenters and attendees. My sincere thanks also go to Philippe Geril, our deus-ex-machina, who has helped us in putting together such an excellent program, as well as for his organizational efforts and input with the Conference. We are also sure that this new start of the series will bode well for their future expansion and they will become the de-facto annual meeting destinations for professionals in the fields of simulation, gaming and industrial simulation in Asia.

Finally we wish you all a pleasant stay in Shanghai and hope you will find time outside the conference to go and explore this wonderful city.

Wenji Mao

Game-On Asia 2010, General Chair

Lode Vermeersch

ASTEC'2010, General Chair

CONTENTS

CONTENTS

SCIENTIFIC PROGRAMME

INVITED PRESENTATIONS

Facial Cloning for Online Interactive Systems

Mingcai Zhou, Yangsheng Wang, Zhijun Du, Zhiyuan Luo
Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
E-mail: {mingcai.zhou, yangsheng.wang, zhijun.du}@ia.ac.cn, zyluo@msn.cn

KEYWORDS
2D+3D AAM, Shape Initialization, Shape Constraint, Expression mapping.

ABSTRACT
In this work, we address the problem of robust real-time facial cloning for online interactive systems. The main contributions of the paper are as follows. For face analysis, we develop an extended 2D+3D Active Appearance Models (AAM) based 3D face and expression tracking framework. Moreover, we introduce a robust shape initialization method to handle agile head motion and a penalty function to avoid unallowable mouth shapes. For face synthesis, we generate online real-time 3D expression by blending the expression bases, and these expression bases are created through an offline expression mapping algorithm.

INTRODUCTION

Facial cloning is useful for many online interactive systems, such as gaming, video conference, and animated avatars for web communication. An online interactive system should work stably and accurately for different illuminations and environments, persons, head motions and face expressions with a real-time performance, and run continuously for long sequences without drift and error accumulation.

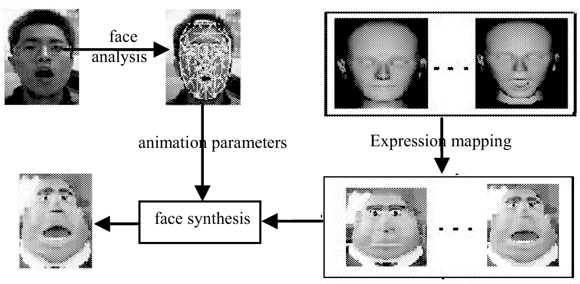


Figure 1. System overview diagram.

The core vision tasks of facial cloning include facial analysis and facial synthesis. Facial analysis is to extract facial expressions and head pose of the user in video stream. It can be considered as a 3D face and facial action tracking problem. Facial synthesis is to generate 3D avatar animation controlled by the tracked facial pose and actions.

We propose a robust and efficient online facial cloning interactive system as shown in Figure 1. Our system is composed of two models: one is face analysis model and the other is face synthesis model. Face analysis model tracks the face and extracts a set of animation control parameters, and then face synthesis model uses these parameters to generate animation of an avatar.

Face Analysis: We extend the principle of 2D+3D AAM for facial analysis. First, we propose a 2D+3D AAM based framework to track 3D face and facial action using three view-based AAMs and a modified 3D face model *Candide* [1]. By using the 3D *Candide* model, we can explicitly recover the 3D pose and facial action. Second, we develop a robust shape initialization method based on local feature matching to provide a good initial shape for AAM fitting. It greatly improves the performance of handling agile head motion. Third, we introduce a penalty function as an additional shape constraint in AAM fitting prevent an unallowable mouth shape which happens frequently in the context of an online tracking.

Face synthesis: For the online interactive system, it requires the 3D facial expression synthesis technique be fast. We present an efficient expression synthesis method whose run-time computation is constant independent of the complexity of the avatar model. The basic idea is to precompute all deformation bases of the target model so that the run-time operation involves only blending together these deformation bases appropriately. These deformation bases are obtained by offline expression mapping method.

STATE OF THE ART

The so-called 2D+3D AAM is firstly proposed by Xiao et al [2]. In [2], only one 2D AAM is used, so it can not handle large angles of head rotation. The shape bases of the 3D shape model they used simultaneously account for shape variability (inter-person variability) and the facial action (intra-person variability) and the two kinds of variability are not independent. Therefore, they have to fitting the shape variation at each frame and can not output facial action parameters explicitly.

Dornaika and Ahlberg [3] proposed to estimate the head pose using a RANdom SAmpling Consensus (RANSAC) technique [4] combined with a texture consistency measure to avoid drifting. However, the RANSAC technique is time-consuming and the texture consistency measurement needs a synthesis step which is also time-consuming.

To generate allowable 2D face shapes, many approaches have been proposed [5, 6]. Li and Ito [5] proposed to control the shape parameters by using discrete shape parameter distribution tables built on training face shape samples. The accuracy of the shape constraint depends on the quantization level of distribution tables.

FACE ANALYSIS

2D Active Appearance Models

The 2D AAM has a compact representation of both shape and appearance as a linear combination of a small number of modes of variation:

$$s(p) = s_0 + \sum_{i=1}^{n} p_i s_i \text{ and } A = A_0 + \sum_{i=1}^{m} \lambda_i A_i$$

where s_0 and A_0 are the mean shape and appearance vectors, $\{s_i\}$ and $\{A_i\}$ are modes of shape and appearance variation, $\{p_i\}$ and $\{\lambda_i\}$ are shape and appearance parameters respectively. The modes of shape and appearance variation are usually built by applying PCA to a set of labelled training images. Figure 2 (a) and (b) show some examples of the shape and appearance variation respectively.

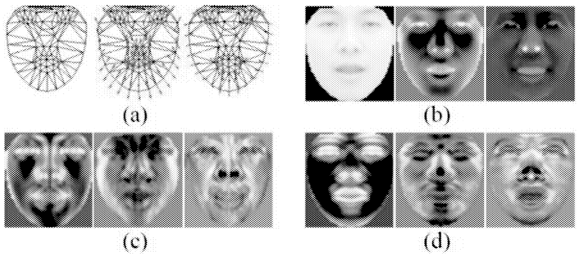


Figure 2. Multi-band AAMs. (a) Mean shape s_0 and the first two shape bases learned by PCA. (b) to (d) are mean appearance and the first two appearance bases of the intensity band, x-direction band and y-direction band respectively.

The goal of 2D AAM fitting is to minimize the difference between the warped-back appearance $I(\mathbf{W}(\mathbf{p}))$ and the synthesized appearance A_λ:

$$E_a(\mathbf{p}, \boldsymbol{\lambda}) = \sum_{\mathbf{x} \in \mathbf{s}_0} \left[A_0(\mathbf{x}) + \sum_{\mathbf{i=1}}^{m} \lambda_i A_i(\mathbf{x}) - I(\mathbf{W}(\mathbf{x}))\right]^2 \quad (1)$$

where $\mathbf{W}(\mathbf{x}, \mathbf{p})$ is a warping function defined to map every pixel $\mathbf{x}$ in the model coordinate to its corresponding image point. The cost function (1) can be efficiently minimized by the inverse compositional parameter update technique [7].

Deformable 3DWireframe Models

In this study, we use the 3D face model *Candide* [1]. The 3D face model is given by the 3D coordinates of vertices $\mathbf{P}_i, i = 1, \cdots, n$, where n is the number of vertices. Therefore, the shape up to a global scale be fully described by the $3n$-vector $\mathbf{g}$ -the concatenat-on of the 3D coordinates of all vertices $\mathbf{P}_i$. The vector $\mathbf{g}$ can be written as:

$$\mathbf{g}(\sigma, \alpha) = \bar{\mathbf{g}} + \mathbf{S}\sigma + \mathbf{A}\alpha$$

where $\bar{\mathbf{g}}$ is the standard shape of the model, and the columns of $\mathbf{S}$ and $\mathbf{A}$ are the Shape and Animation Units respectively, and thus the vectors σ and α contain the shape and animation parameters. The term $\mathbf{S}\sigma$ accounts for inter-person variability while $\mathbf{A}\alpha$ accounts for facial animation (intra-person variability). Moreover, we assume that the two kinds of variability are independent.

2D+3D AAMs

The main idea of 2D+3D AAMs is to impose that there exist legitimate values of rigid and non-rigid deformation parameters such that the 2D projected 3D shape consists with the 2D shape of the 2D AAM. This constraint can be written as:

$$\min Em = \min \left\| \mathbf{s}'(\mathbf{p}) - P(Q(\mathbf{g}'(\sigma, \alpha))) \right\|^2 \quad (2)$$

where $\mathbf{g}(\sigma, \alpha)$ is the 3D *Candide* model, $Q(\mathbf{x})$ is the rigid transform with rotation matrix $\mathbf{R}$ and translation vector $\mathbf{T}$, $\mathbf{P}$ is the full-perspective projection of the 3D vector $Q(\mathbf{x})$ in the image plane. The superscript "$'$" for $\mathbf{s}$ and $\mathbf{g}$ denote that we only use partial vertices in the vector as constrained points because there is not always one-to-one correspondence between the 2D shape and the 3D shape. Figure 4 (a) shows the constrained vertices (marked by the points) on the 3D model.

There are two advantage of using the *Candide* model. The first advantage is fitting speed. We fit the shape parameters σ just at the first frame and then fix them for the following frames. The second advantage is that we can recover the facial action (i.e. α) explicitly.

The goal of the so-called 2D+3D AAM fitting is to minimize:

$$E = E_a + w_m E_m, \quad (3)$$

where w_m controls the strength of 3D shape constraint.

By experiments, we found that $w_m = 0.1$ is appropriate for most cases. The cost function (3) can be efficiently minimized based on inverse compositional technique.

2D+3D AAMs based Face tracking

In our system, we extend the basic 2D+3D AAM technique in the following aspects to make the tracking more stable.

Multi-Band AAMs with Edge Structure. We adopt the multi-band appearance model [8] to improve the tracker's generalizability. In our system, the appearance is a concatenation of three texture bands' values: the intensity, x-direction gradient strength and y-direction gradient strength. Figure 2 shows the leading shape and appearance bases of our multi-band AAMs.

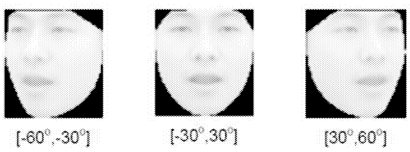


Figure 3. View-based AAMs. Here just shows the mean appearances of the intensity band for each view.

View-Based AAMs. To handle large angles of head rotation, we adopt a view-based approach [9]. Three 2D AAMs are trained for view ranges [−60°, −30°], [−30°, 30°] and [30°, 60°] respectively. Figure 3 shows the mean appearances of the intensity band for each view.

Extended 3D *Candide* Model. To handle large angles of head rotation, we extend the original 3D *Candide* model by adding some vertices and corresponding meshes (marked by the rectangles in Figure 4 (b) and (c)) on the cheek. Figure 4 (b) and (c) show the left and right view of the extended 3D *Candide* model with the constrained points respectively.

Tracking Initialization and Recovery. In our system, we judge the tracking to be lost if the appearance reconstruction error (Equation 1) exceeds a pre-defined threshold. To initialize or recovery the tracking, we first do face detection [10], then fit AAM using the detected face bounding box as initialization.

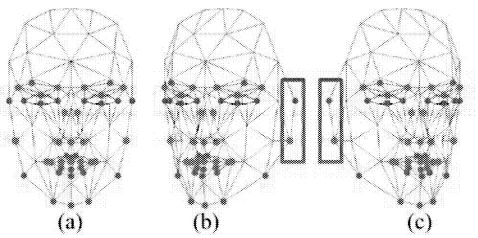


Figure 4. 3D Wireframe Model. (a) The original 3D Candide model with the constrained points (the red points). (b) and (c) are the left and right view of the extended 3D Candide model with the constrained points.

Robust Shape Initialization Based on Face Motion Direction

Good initial parameters are crucial to the success of AAM fitting. We develop a fast shape initialization algorithm based on local feature matching in the context of an online tracking. We firstly select some interesting feature points by a corner detector [11] and some semantic points such as the eyes' corners at frame $t-1$. Then, at frame t, we adopt correlation-based local search to find the matched feature points to frame $t-1$. Next, we reject the outliers by the main direction of the face motion. Finally the shape parameters $\mathbf{p}_t$ are estimated to fit to the remained points.

As shown in Figure 5, for different face motions, there usually exists a main motion direction. Those feature points whose moving directions are not consistent with the main direction are most likely to be outliers. Denote θ as the moving angle of a feature point and $\theta \in [0,2\pi]$, we estimate the main motion direction based on the angle θ's histogram, as shown in Figure 5 (d). The main motion's angle Θ corresponds to the bin with maximum points. Given a threshold $\Delta\Theta$, we filer out those points whose moving angles are not consistent with the main direction: $\theta \notin [\Theta - \Delta\Theta, \Theta + \Delta\Theta]$. In our experiments, $\Delta\Theta$ is set as $\pi/9$. Figure 5 (c) shows the filtered results by the main direction. As we can see, those miss-matched points are rejected effectively. Comparing with RANSAC [4], our motion direction filter is more efficient with comparable results.

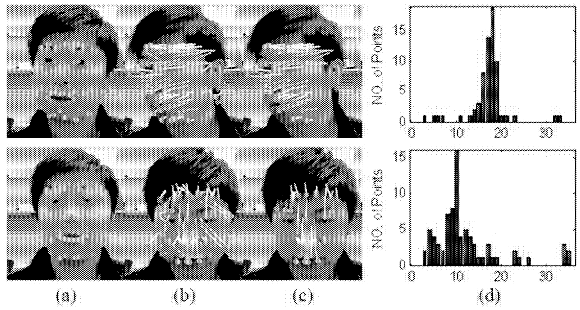


Figure 5. Reject the matching outliers by main direction filter. (a) Selected feature points at frame $t-1$. (b) Matched feature points at frame t. The lines show the moving directions of the points. (c) Remained feature points after main direction filter. (d) The histogram of the moving angle. Here we discretize the angle's range $[0,2\pi]$ to 36 bins and add an additional bin for the fixed points.

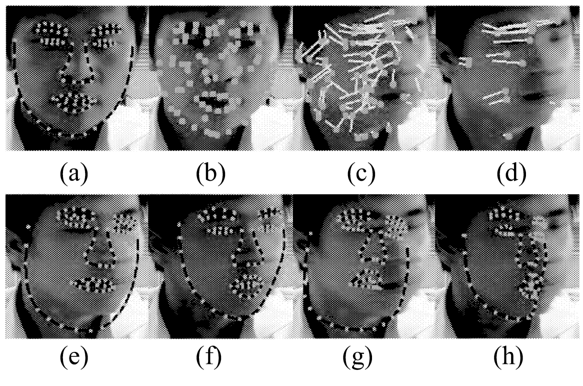


Figure 6. Comparisons of different shape initialization methods. (a) The shape of frame $t-1$. (b) Selected feature points at frame $t-1$. (c) Matched feature points at frame t. (d) Remained feature points after main direction filter. (e). The initial shape of frame t. (f) the result shape of frame t. (g) The initial shape estimated from all the matched feature points without main direction filter. (h) The result shape using (g) as initial shape.

Suppose after filtering by the main motion direction, M matched points $\{\mathbf{z}_i\}_{i=1}^{M}$ are remained, we estimate the initial shape parameters $\mathbf{p}_0$ to fit to these points, that is, to minimize the following cost function:

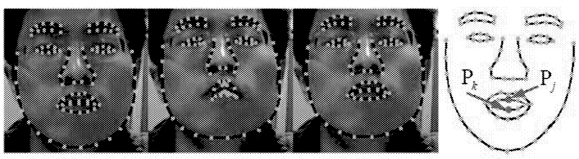
$$p_0 = \min_{p} \sum_{i=1}^{M} - w_i \rho(\| \sum_{j=1}^{3} c_{ij} W(x_{ij}; p) - z_i \|^2, r) \qquad (4)$$

where the weight $w_i = \cos(\theta_i - \Theta)$ is set as the consistency with the main motion direction, $\rho(\cdot, r)$ is a m-estimator as in [12], and $\sum_{j=1}^{3} c_{ij} W(\mathbf{x}_{ij}; \mathbf{p})$ is the estimated position of the point i given the shape parameters $\mathbf{p}$, $\{\mathbf{x}_{ij}\}_{j=1}^{3}$ are the vertex coordinates of the triangle containing the point i in the model coordinate, $\{c_{ij}\}_{j=1}^{3}$ are the triangle coordinates of the point i. We find the optimal $\mathbf{p}_0$ by Gauss-Newton algorithm [13].

Figure 6 compares the results of different initialization methods, as we can see, main direction filter generates a much better initial shape which encourages the correct alignment.

Penalty Function Based Shape Constraint

The Shape model of a 2D AAM assumes that the shape instances satisfy a multi-dimensional Gaussian distribution. Therefore, any shape instance within a hyperellipsoid whose radius is proportional to the corresponding eigenvalue is assumed to be a allowable shape by the shape model. However, because of the non-Gaussian nature of the sample distribution, this assumption may be not true. Our experimental results show that unallowable shapes can be generated in some cases even if it is within the hyperellipsoid. One example is shown in figure 7 (b), where the shape of the mouth is an unlikely one. This kind of shape often happens when a user looks up to the web camera and so an upward view is generated.



(a) (b) (c) (d)

Figure 7. Comparison of with and without the shape constraint. (a) The shape of frame $t-1$. (b) The result shape of frame t without shape constraint. (c) The result with shape constraint. (d) The two constrained points on lower lip and upper lip respective.

To address this problem, several approaches have been proposed by many researchers [5, 6]. The method proposed by Li and Ito [5] need to build many discrete shape parameter distribution tables during the training stage and the shape constraint is not accurate for the low level of the quantization. In our study, the most frequent case is the unlikely mouth shape shown in Figure 7 (b). Therefore, we propose a more direct constraint on the mouth shape. That is, we impose that the points belong to lower lip must be below those points belong to upper lip. In fact, we can just impose this constraint on a point $\mathbf{P}_k$ belong to lower lip and another point $\mathbf{P}_j$ belong to upper lip, as shown in figure 7 (d).

Mathematically, the shape constraint can be written as:

$$E_s = [\max(0, -(\mathbf{s}_0^{yk} + \sum_{i=1}^{n} p_i \mathbf{s}_i^{yk} - (\mathbf{s}_0^{yj} + \sum_{i=1}^{n} p_i \mathbf{s}_i^{yj}))]^2$$

where $\mathbf{s}_o^{yk}$ is the y coordinate of point $\mathbf{P}_k$ in mean shape $\mathbf{s}_0$. Other terms are similar. Note that here we don't consider the four additional shape bases representing global transformation. Comparing with the method proposed by Li and Ito [5], our method does not need the training process and the constraint is more intuitionistic and direct.

Considering all of the constraints, the final cost function is:

$$E = E_a + w_m E_m + w_s E_s, \qquad (5)$$

where w_s controls the strength of online appearance model constraint and allowable shape constraint respectively. By experiments, we found that $w_s = 0.01$ are appropriate for most cases. The cost function (5) can be efficiently minimized based on inverse compositional technique [7].

FACE SYNTHESIS

As we have described, the tracking model outputs head pose control parameters and expression control parameters, we will discuss how to use theses parameters to animate an avatar. Many facial animation techniques have been proposed, such as expression cloning technique presented by Noh and Neumann[14], blend shape interpolation method[15], MPEG based facial animation[16] and so on. The avatar animation in our system is controlled by the tracked parameters, so these techniques cannot be directly used. In order to generate real-time interactive animation, the proposed animation technique should be very fast. According to these requirements, we propose an animation technique which is similar to the blend shape interpolation method, in the online tracking stage, we synthesize expressions by interpolating the basic expressions controlled by the expression parameters, and we generate these basic expressions offline when a new avatar is added to the library.

Online Animation Generation

Table1. Relationship between expression control parameters and facial parts

Facial parts	Number of parameters	Description
Entire face	3	three rotattion angles
Left eyebrow	2	raise / squeeze left eyebrow
Right eyebrow	2	raise / squeeze right eyebrow
Left eye	1	close left eye

Right eye	1	close right eye
Mouth	5	mouth in five different expressions
Nose	0	no special process

The avatar animation includes head movement and local expression deformations, both of them are controlled by the actor. The avatar head pose can be directly derived from the tracked pose control parameters R and T, and the avatar local expression deformation is determined by the tracked expression control parameters α. Each facial part is controlled by one or more expression parameters, table 1 gives detailed description. Expression parameter α_i is a scalar, and $\alpha_i \in [0,1]$, it controls the intensity of movement. Take the left eye for example, in our system α_1 controls its state. $\alpha_1 = 0$ indicate it is open and totally closed as $\alpha_1 = 1$. There are 5 parameters to control the mouth movement, and each parameter corresponds to one expression. Nose does not show significant movement during animation, so there is no parameter to control it.

For each expression control parameter α_i, we create a corresponding basic expression V_i, Figure 8 illustrates some basic expressions in our system. Note that V_i is the expression when $\alpha_i = 1$, $\alpha_j = 0, j \neq i$, namely it only deform its local facial region, while other facial parts remain the neutral state.

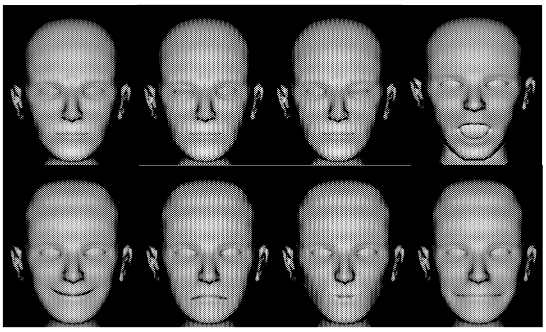


Figure 8. Some basic expression examples

For each frame, expression control parameters α_i is calculated through facial expression analysis model, the avatar's 3D expression is obtained by interpolation.

$$V = \sum_{i=0}^{n} \alpha_i V_i \ , \ 0 \le \alpha_i \le 1;$$

Avatar expression V is then combined with head pose to generate the final animation which is similar to the actor's action.

Offline Basic Expression Generation

In the previous section, we precompute all the basic expressions of the target avatar so that the run-time operation involves only blending together these bases appropriately. When a target avatar is loaded, we should first synthesize its basic expressions, we use expression retargeting technique to map the expression of a source model to the target avatar.

We first ask an artist to create a source neutral model and its 12 basic expressions as Figure 8 shows, and then we map these basic expressions to the target avatar. We require source model and target avatar be in the neutral expression, the expression mapping process is achieved by two stages: key-point motion transfer and motion interpolation.

Key-point motion transfer： We first label some key points on the models to manually create a small set of correspondences between these two surfaces, and then transfer the key point's motion vector on the source model to the target model. The deformation on the source surface can not simply be transferred to a target model without adjusting the direction and scale of each motion vector, because facial proportions and geometry vary between models. Based on these key points, Radial Basis Function[17] can be used to learn the mapping function from the neutral expression of the source surface $\mathbf{x}_s = (x_s, y_s, z_s)$ to the neutral expression of the target surface $\mathbf{x}_t = (x_t, y_t, z_t)$, such that $\mathbf{x}_t = f(\mathbf{x}_s)$. Given a small deformation for a source point $\boldsymbol{\delta}\mathbf{x}_s = (\delta x_s, \delta y_s, \delta z_s)$, the deformation $\boldsymbol{\delta}\mathbf{x}_t$ of its corresponding target point $\mathbf{x}_t$ is computed by the formula $\boldsymbol{\delta}\mathbf{x}_t = \mathbf{M} \cdot \boldsymbol{\delta}\mathbf{x}_s$, while $\mathbf{M}$ is the Jacobian matrix at $\mathbf{x}_s$ numerically.

Motion interpolation: Given the deformation vector of the key points on the target surface for each basic expression, this step deform the remaining vertices on the surface by linearly interpolating the movement of the key points using barycentric coordinates. First, the system generates a triangular mesh based on the key points. Second, spherical parameterization [18] is used to help determine each vertex's barycentric coordinate. Then the deformation vector of the remaining vertices is interpolated accordingly.

EXPERIMENTS

Face Analysis Performance

To verify the effectiveness of our face analysis method, the algorithm has been tested on the videos of three persons. The videos are all downloaded from web sites that are taken under uncontrolled environments. Figure 9 shows some frames of the videos.

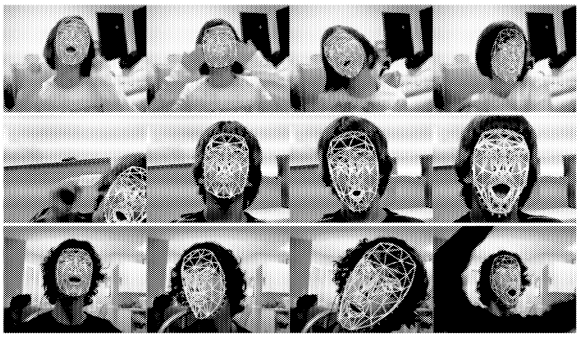


Figure 9. Tracking results of the videos.

In our experiments, we train a multi-view AAM model with multi-bands as explained in previous section. For each view's AAM, the shape model contains 8 shape bases, the appearance model contains 40 appearance bases and the size of the appearance patch is 52×58.

Figure 9 shows the typical face tracking results of the three videos. Our proposed algorithm accurately localizes the facial components under the conditions with illumination changes, large expression and pose variations. Our face analysis algorithm is realtime. On a Pentium-4 3.0G computer, the algorithm's speed is about 60 fps for the video with 320×240 resolution.

Online Facial Interactive System

We implement a fully automatic online facial interactive system based on our face analysis and synthesis algorithms. Any common web camera can be used in the system. On a Pentium-4 3.0G computer, the whole system's speed is about 35 fps for the video with 320×240 resolution.

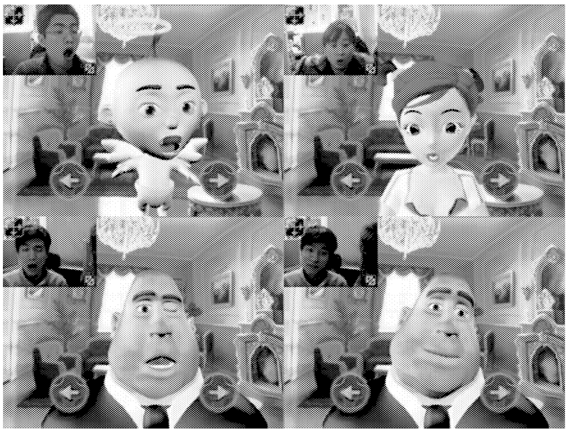


Figure 10. Some facial cloning results.

Figure 10 shows some facial cloning results. As we can see, our system accurately extracts head pose and facial expressions and then retargets the pose and expressions on to the avatar.

CONCLUSIONS

In this paper, we proposed a solution to solve facial cloning for online interactive system. For face analysis, we proposed an effective and efficient 3D face and facial action tracking algorithm based on 2D view-based AAMs and a modified 3D face model. The robustness to fast motion is greatly improved by robust shape initialization. A penalty function is introduced to effectively prevent unallowable mouth shapes. For face synthesis, we proposed an efficient expression synthesis method whose run-time computation is constant independent of the complexity of the avatar model.

FUTURE WORK

Currently our online interactive system can not extract subtle facial actions, such as wrinkles. We plan to use the pattern recognition method to extract the subtle actions.

REFERENCES

[1] J. Ahlberg. Candide-3 - an updated parameterised face. Technical report, 2001.

[2] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2d+3d active appearance models. In *CVPR (2)*, pages 535–542, 2004.

[3] F. Dornaika and J. Ahlberg. Face model adaptation using robust matching and active appearance models. In *WACV*, pages 3–7, 2002.

[4] C.-P. Lu, G. D. Hager, and E. Mjolsness. Fast and globally convergent pose estimation from video images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(6):610–622, 2000.

[5] Y. Li and W. Ito. Shape parameter optimization for adaboosted active shape model. In *ICCV*, pages 251–258, 2005.

[6] L. Gu, E. P. Xing, and T. Kanade. Learning gmrf structures for spatial priors. In *CVPR*, 2007.

[7] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

[8] M. Zhou, Y. Wang, X. Feng, and X. Wang. A robust texture preprocessing for aam. In *CSSE (2)*, pages 919–922, 2008.

[9] T. F. Cootes, G. V. Wheeler, K. N. Walker, and C. J. Taylor. View-based active appearance models. *Image Vision Comput.*, 20(9-10):657–664, 2002.

[10] P. A. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[11] J. Shi and C. Tomasi. Good features to track. In *CVPR*, Seattle, June 1994.

[12] J. Pilet, V. Lepetit, and P. Fua. Real-time non-rigid surface detection. CVPR05.

[13] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, USA, August 1999.

[14] J.Y. Noh and U. Neumann. Expression cloning. In Proceeding of SIGGRAPH 01, 277-288, 2001.

[15] F.Pighin, R. Szeliski and D.H. Salesin. Modeling and animating realistic faces from images. International Journal of Computer Vision, 50(2):143-169,2002.

[16] G. Abrantes and F. Petreira. MPEG-4 facial animation technology: survey, implement and results. IEEE Transactions on Circuit and System for Video Technology, 9(2):290-305, 1999.

[17] F. Ulgen. A step towards universal facial animation via volume morphing. In Proceeding of 6th IEEE

international Workshop on Robot and Human Communication, 358-363,1997.
[18] E. Praun and H. Hoppe. Spherical parameterization and remeshing. In proceeding of SIGGRAPH 03, 340-350, 2003.

BIOGRAPHY

MINGCAI ZHOU received the BS and MS degree in Mechanical and Electrical Engineering from Huazhong University of Science and Technology, China in 2001 and 2004 respectively. He is currently a PhD candidate at Institute of Automation, Chinese Academy of Sciences. His research interests include computer vision, image processing, human computer interaction, with special attention to face alignment and tracking.

YANGSHENG WANG received the BS and MS degree in Information Technology, and PhD in Computer Vision, from Huazhong University of Science and Technology, China in 1982, 1984 and 1989 respectively. He was a senior academic visitor in Meta Machines Ltd, Oxford of UK for one year in 1987 and worked in the Department of Artificial Intelligence, Edinburgh University as a Postdoctoral Fellow in 1989. He has been with Institute of Automation, Chinese Academy of Science of China as a Professor since 2000.

ZHIJUN DU is currently a PhD student. He studies computer science at the Institute of Automation, Chinese Academy of Sciences, Beijing. His research interests mainly include face expression analysis, face modelling and animation.

ZHIYUAN LUO received the MS degree in 3D animation from Peking University and the BS degree in Information and Computing Science from Sun Yat-Sen University. He is currently a researcher at the Institute of Automation, Chinese Academy of Sciences. Before joining the Institute of Automation, he worked as a member of technical staff at Beijing ShengKai Technology, which is a branch of ShengDa Internet Development, focusing on new technology and methods of interactive display. His research interests include 3D face modeling and facial animation, new technology and methods of interactive display and medical visualization.

Modeling Rich Characters in Interactive Narrative Games

Mei Si, Stacy C. Marsella
Institute for Creative Technologies
University of Southern California
Email: {meisi,marsella}@ict.usc.edu

KEYWORDS

Virtual Character, Interactive Narrative, Theory of Mind, Social Norm, Emotion.

INTRODUCTION

Computing technologies have advanced rapidly over the past decade. Faster machines, better graphics, and more advanced algorithms become available every year. Moreover, the evolution of internet technology and the increasing accessibility of computing resources and mobile devices allow computing technologies to go beyond business and scientific computing, and become an important means for providing entertainment and facilitating communication. These advances have helped to enable a new form of media – interactive narrative games. Interactive narrative games allow a user to play a role in a story and interact with other characters driven by AI agents. The user's choices affect the unfolding of the story.

Because of the support of user interactivity and the use of computer simulated virtual environments, interactive narrative games are closely related to video games. In fact, the rapid growth of interest in interactive narrative games is in part motivated by the explosion of computer-based games in recent years. Compared to more traditional forms of video games, such as arcade games, action games, and even role playing games, interactive narrative games emphasize more of the social and narrative aspects of the experience. Story, of course, is a central part of the human experience both as entertainment and as a powerful tool for providing pedagogy. We watch movies, read novels and tell stories. Interactive narrative games provide an experience that integrates user agency with the engaging power of narrative.

Interactive narrative games have been recognized as a promising tool for providing both pedagogy and entertainment. They have been proposed for a range of training applications, e.g. [13, 20, 26, 35, 22] as well as entertainment applications, e.g. [10, 23, 3, 12, 11, 36].

In this paper, we discuss the design desiderata for interactive narrative games, and in particular for creating the virtual characters in interactive narrative games. We argue that a rich model of characters that are well-motivated, socially aware and have a "Theory of Mind" is needed. We discuss the state of the art work on modeling virtual characters. In particular, we present the approaches taken in Thespian [27, 26, 28, 29, 31, 30] – a decision-theoretic multi-agent framework for interactive narratives.

DESIDERATA FOR INTERACTIVE NARRATIVE GAMES AND VIRTUAL CHARACTERS

In media that involves narratives, such as movies, dramas and interactive narrative games, the coherence of the narrative is a basic design goal. Narrative, which is typically defined as "the semiotic representation of a series of events which are meaningfully connected in a temporal and causal way" [18] (Note that "Narrative" can be defined and used in more general ways. In this paper, we use the term "coherent narrative" or "coherence of narrative" to refer to this definition), has been shown as an important way for people to organize and make sense of their experience [2, 16, 17]. If the experience does not make sense to the viewer/user, the author cannot possibly reach his/her other goals.

Coherent narrative by itself does not necessarily lead to a dramatic or inspiring experience because it merely requires the causal and temporal relationships between events to be understandable to the reader/user. The author of interactive narratives often wants to create cognitive or affective effects in the user through the interactive experience. For example, many interactive narratives for training are designed to help the user practice social and cognitive skills, such as (social) problem solving [20], negotiation [38] and coordination skills [35] when the user is at certain affective states, e.g. highly stressed. More generally, interactive narratives have often depended on triggering the user's emotional responses to keep the user engaged and set up the environment for the user to learn or get entertained. For example, in FearNot! [20], which is targeted at helping the learner deal with school bullying, empathic responses for the victim are triggered by letting the user talk to a child character who is the victim of a school bully. In both Façade [15] and Mimesis [23], the systems have a major goal of creating a dramatic tension arc, i.e. a slow increase in tension followed by a release. And of course, this is not unique to interactive narrative. Narrative forms in general seek to create cognitive or affective results. Tan, for example, describes films as "emotion machines" [37].

In discussing traditional narratives, Egri has suggested a central, key role for rich, fully fleshed out characters for narrative design. He argued that such characters are critical not only to narrative but also critical as key aspect to the process of creating narrative – that rich characters achieve autonomy in the writer's mind and can thereby serve as inspiration to the author [6].

The richness of character development is evidenced in the works of Shakespeare. The play Othello, for example, gives us a sense of the richness of character that an author may seek. In this play, Iago hates Othello and seeks his downfall. He hatches a plan to plant evidence that will lead Othello to the false inference that his wife has cheated on him. Iago believes that this false inference will lead Othello to kill his wife and consequently destroy himself. Here the richness of character can be observed. The characters have beliefs about others including how others think - they possess what in psychology and philosophy is called a "Theory of Mind". They have motivations and emotions. Finally, they understand the social structures and roles of which they are a part, such as marriage and spouse, the social norms associated with those roles and the consequences of violating them. In other words, the characters are well-motivated and socially aware.

Similarly, in interactive narrative games, we need to build characters that are well-motivated, have a "Theory of Mind", understand social norms and have emotions. These capacities in characters are not only important for creating a dramatic effect, but also necessary for allowing the user to understand the characters as they enable the characters to behave human-like.

CHALLENGES IN CREATING VIRTUAL CHARACTERS

The design of rich, human-like characters faces many challenges. This section discusses one of the key challenges – balancing the design of characters and the design of events.

Story is composed of characters and events. Any effects the author intends to reach are created by the design of the characters and the design of the events. Though characters and events are closely related, their designs are often in conflict with each other. Overemphasizing the design of events will result in having over-simplified or broken characters because the characters do not possess their own motivations and personalities, but are designed only to support the events. On the other hand, even with very rich characters, plot is still important. If all the consideration is given to craft characters, the overall story may lose its structure, and becomes a trivial story in which nothing happens [5]. An extreme example is a chat room, where the user interacts with real people but hardly has any structured experience.

In general, a balance has to be reached between the design of the characters and the design of the events. They serve as constraints to each other – when designing characters the author needs to think about how the characters' behaviors can be used to achieve the plot design, and when designing events the author needs to make sure that the characters are not broken or losing their distinct personalities for acting out the events.

THE STATE OF THE ART

There is a considerable body of work in the field of artificial intelligence and multi-agent systems on building interactive narrative games. In this section, we briefly review how virtual characters are modeled. This review compares how different interactive narrative frameworks create coherent narratives in the face of user interaction and how social normative behaviors and emotions are modeled in the characters.

Create Coherent Narrative

The coherence of narrative requires the causal and temporal relationships among events to be interpretable to the user, which usually implies that characters need to have consistent and human-like motivations. Most of the existing works on interactive narratives can be viewed as either adapting a plot-centric design or a character-centric design for creating coherent narrative.

Plot-centric designs emphasize the design of the events in the story, and the characters' actions are driven by the development of the plots. For example, In Façade [15], the story is organized around hand-authored dramatic beats. Based on a desired global plot arc, the drama manager chooses the next beat that is suitable to the context and whose dramatic value best matches the arc. In Mimesis [23], the authoring framework constructs story plans, which are the ideal linear narrative that the user should be told. When the user's action deviates from the story plan, the system either replans or prevents the user's action from being effective. I-storytelling [3] system plans over a hierarchical tasks network (HTN) to realize interactive narratives.

In contrast, character-centric approaches for interactive narrative emphasize the design of individually plausible characters. The story emerges from the user's interaction with the characters. For example, FearNot! [20] uses planning based approach for modeling the characters' behaviors. It has explicit representations of characters' personalities and motivations, which affect the individual character's plan construction process. In MRE [35] and SASO [38], there is an extensive dialogue management subsystem in each character that incorporates explicit rules for dialogues. The agents have plans governing the coherence of their behaviors which take their personalities into account.

Most of the contemporary interactive narrative frameworks provide systematic support of either the design of character or plot structure, but rarely both. In these frameworks, it is usually up to the human author to ensure the design of the other component, which often turns out to be a significant undertaking. For example, it is extremely time consuming if not impossible to manually check whether the characters are well-motivated in all the possible paths through the story.

Model Social Normative Behavior

Social norms are commonly believed rules in social interaction. These rules serve as a guide for human behavior, and as the basis for their beliefs and expectations about others. Though norms are commonly followed, the tendency to follow norms is regulated by other factors, such as more pressing, personal goals.

In interactive narrative games, norm-following/violating behavior is often not explicitly modeled. Rather, they are modeled conjointly with characters' other behaviors. For example, in Façade [15], norms are encoded in the design of the beats and the beat selection process, i.e. the pre- and post-conditions of the beats. In I-storytelling [4], characters' behaviors including norm following behaviors are modeled using hierarchical task network (HTN) plans. In MRE [35] and SASO [38], the dialogue management subsystem incorporates explicit rules for normative behaviors, specifically conversational norms. The priorities of these rules are adjusted by agent authors to fit the characters' profiles.

Model Emotions

In modeling emotion, cognitive appraisal theories have had an increasing impact on the design of virtual characters. Appraisal theories are a class of leading psychological theories for emotion. Appraisal theories argue that a person's subjective assessment of their relationship to the environment, the person-environment relation, determines the person's emotional responses [24, 33, 19, 34, 25, 7]. For example, an event that is incongruent with the person's motivations and is caused by others may lead to anger responses; on the other hand, if the event is caused by the person himself/herself, the person will feel guilty or regret [24].

Similar to how social normative behaviors are modeled, many interactive narrative frameworks do not have an explicit model for the characters' emotions. Here we briefly review frameworks that model emotion explicitly. FearNot! [1] deployed the OCC model of emotion over its plan based agents. EMA [8], which is the emotion subsystem in MRE and SASO, follows the Smith and Lazarus theoretical model of appraisal [34]. EMA [8] defines appraisal processes as operations over a uniform plan-based representation, termed a causal interpretation, of the agent's goals and how events impact those goals. Cognitive processes maintain the causal interpretation and appraisal processes leverage this uniform representation to generate appraisal.

THE THESPIAN FRAMEWORK

This section presents a unique framework – Thespian [27, 26, 28, 29, 31, 30] – for interactive narrative design, which utilizes autonomous agents for well-motivated and socially aware characters, and multi-agent coordination to realize story plots. Thespian is able to create both rich characters and manage the development of the story during the interaction according to author specified plot design goals.

Thespian has been applied to authoring dozens of virtual characters in more than thirty interactive narratives. The first interactive narrative to incorporate Thespian is the Mission Environment of the Tactical Language Training System (TLTS) [9], which is aimed at providing rapid language and culture training. Thespian has also been used to model fables such as "the Little Red Riding Hood story" and the Fisherman and his wife. We will use "the Little Red Riding Hood" story as an example to motivate the discussion throughout this section.

Overview of Thespian

Egri Lajos, who is famous for his teaching of creative writing, has strongly argued for the importance of characters in narratives [6]. His view of narrative – of rich, well motivated, autonomous characters as a creative spark to the author, but that are nevertheless constrained by the author's goals for the plot – serves as inspiration to the approach taken in Thespian.

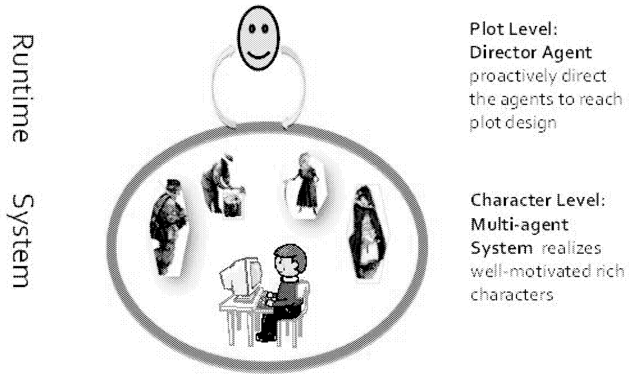


Figure 1: Two-layer Runtime System

Thespian uses a two-layer system for simulating interactive narrative, as shown in Figure 1. The first layer – the multi-agent system – is built based on PsychSim [14, 21], a multi-agent modeling tool for social simulation based on Partially Observable Markov Decision Problems (POMDPs) [32]. In Thespian, autonomous POMDP based agents are used for modeling each character in the story. These agents' behaviors are well-motivated and socially aware. They respond to the user based on both their motivations and the status of the interaction. To guide the development of the story, and therefore give the author control of the plot development, the agents are adaptively fine tuned during the interaction by a director agent, based on the author's plot design and the agents' prior interaction history with the user. In addition to the two-layer simulation system for interactive narrative, Thespian also contains off-line authoring processes to facilitate the author in designing the characters.

Next we will look at the model of the characters in more detail.

Character Modeling in Thespian

Create Coherent Narrative

Thespian creates coherent narratives by modeling characters with consistent and human-like motivations. Thus, the characters act like people and can therefore be interpreted in similar ways. Thespian contains a director agent for managing the development of the story, e.g. for creating certain plot structure during user interaction. When there is a conflict between the design of characters and the design of events, Thespian gives priority to keeping characters' motivations consistent.

In Thespian, each character in the story is controlled by a decision-theoretic goal-based agent. Each agent is composed of its state, actions, dynamics, goals, beliefs and policy.

An agent's state keeps track of the agent's physical and social status in the story. State is defined by a set of state features, such as degree of hunger, being alive, and degree of affinity with another character. The values of state features can be changed by both the agent's own actions, e.g. eat, and other characters' actions, e.g. being killed. Action dynamics define how the values of state features are affected by actions.

The character's motivations are encoded as the agent's goals. Each agent has multiple and potentially competing goals, e.g. keeping safe vs. keeping others safe, that can have different relative importance or preferences. For example, the wolf character can have goals of keeping safe and preventing itself from starving, with the former goal ten times more important than the latter. If the importance of the wolf's goals is the other way around, i.e. it is much more important for the wolf to not feel hungry than to keep himself alive, the wolf will try to eat people regardless of the situation.

Thespian agents have recursive beliefs about self and others, e.g. my belief about your belief about my goals, which forms a "Theory of Mind". The "Theory of Mind" capacity enables the agents to reason about others when making decisions, and thus makes them "social characters". When deciding what to do, a bounded lookahead policy is used by the agents. They project limited steps into the future, considering not only their own actions, but also other characters' responses using their mental models of other characters, and their responses in return. The agents choose the action that receives the highest expected reward to proceed. Thus, they act both true to their motivations and in reaction to the status of the world. For example, in the scenario shown in Figure 2, the wolf will react to Red differently depending on whether there is somebody else close by, and who is that. The wolf will choose different actions when the hunter is near and when the woodcutter is near, because the wolf has different mental models about these two characters.

The user is also modeled using a Thespian agent based on the character whom the user takes the role of. This model allows other agents to form mental models about the user and the director agent to reason about the user's beliefs and experience.

As part of the effort for modeling socially aware characters, Thespian models social normative behaviors and emotions. Next, we look at these models in more detail.

Model Social Normative Behaviors

Different from most interactive narrative frameworks, Thespian explicitly models norms in face to face communication using a domain-independent model built within

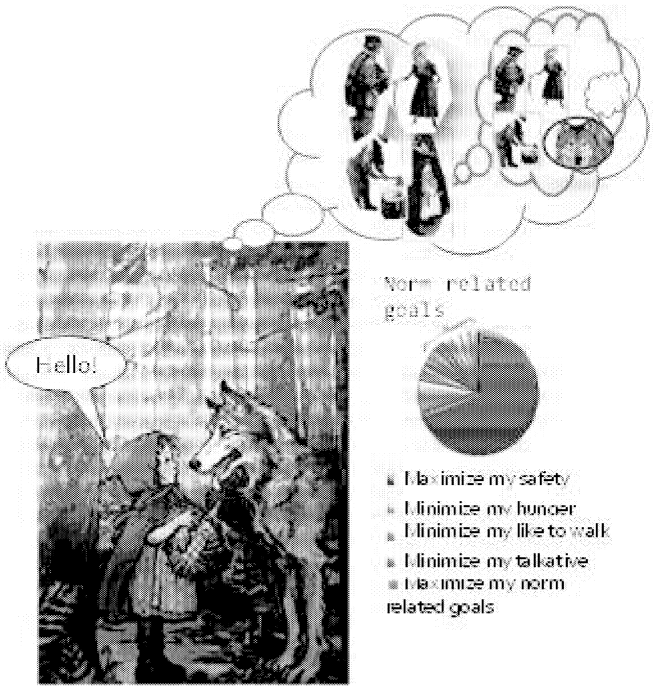


Figure 2: An Example of Character's Motivations and Theory of Mind

a decision-theoretic context [28]. More specifically, three norms are modeled: making relevant responses, following natural turn-taking patterns, and having appropriate conversational flow emerged. The computational model for norms consists of goals that motivate characters to behave socially appropriately, state features that keep track of the status of conversation, and action dynamics for updating these state features. Thespian agents representing characters are given explicit goals of following norms in addition to their other goals. Thus, the characters are allowed to reason about the effect of following or violating norms and achieving or sacrificing their other goals using a unified decision-theoretic framework. By default, Thespian agents act following norms, unless they have other more pressing goals. For example, the wolf will follow norms and greet Red if the hunter is not close by. If the wolf sees the hunter approaching, it will run away and ignore the norm.

Model Emotions
Thespian's model of emotions is based on Smith & Lazarus' appraisal theory [34]. Smith & Lazarus described two types of appraisal: primary appraisal and secondary appraisal. Primary appraisal concerns whether and how the event is relevant to the person. Secondary appraisal evaluates the person's potential for coping with the event. The result of the evaluation will be taken into account by the next primary appraisal process, and thus form an appraisal-coping-reappraisal circle in the agent's cognitive/emotion generation process.

Thespian models five key appraisal dimensions: motivational relevance, motivational congruence, accountability, control and novelty. Upon observing a new event – an action performed by an agent or the human user, each agent appraises the situation along these dimensions based on its beliefs and past expectations. When the agent makes its next decision, its coping potential is reevaluated. It also forms new beliefs and expectations. This updated information will be used for the agent's next appraisal process.

Thespian agents have mental models of other agents; they not only have emotional responses to the environment but can also form expectations of other agents' emotions. For instance, agent A can use its beliefs about agent B to evaluate the motivational relevance and novelty of an event to agent B. The result may be totally different from the appraisal performed from its own perspective. This allows us to create richer characters.

DISCUSSION AND FUTURE WORK

Thespian is a multi-agent framework for authoring and simulating interactive narratives. Its two-layer runtime system is capable of both creating rich characters and managing the development of the story during the interaction according to author specified plot design goals. Most character-centric approaches for interactive narrative can be viewed as only having the first layer of Thespian's runtime system. Therefore, it is usually hard for them to effectively control the development of the story in the face of user interactions. On the other hand, in frameworks that use plot-centric approaches, because a sophisticated character model is missing, the system cannot reason about how a well-motivated character should behave along different paths through the story. As a result, the author has to either sacrifice the richness of characters or spend extensive effort to define their behaviors.

Our future work is planned in several directions. First, we plan to further enrich our models of emotions and social normative behaviors. Secondly, to create the experience of "presence" is one of the ultimate goals of designing virtual environments. Despite that, many fundamental questions about this experience remain open, such as how to measure the user's experience without interrupting it. We are interested in studying how presence can be created and measured in interactive narrative games. Finally, social computing as a new paradigm of computing and technology development, has received increasing attention in recent years. Social computing refers to the computational facilitation of social studies and human social dynamics as well as the design and use of information and communication technologies that consider social context [39]. We want to extend our work on modeling virtual characters to facilitate computer-mediated communications, such as virtual conferences and online chat rooms.

SUMMARY

Interactive narrative game is a new emerging field that has received increasing attention in recent years. Interactive narrative games provide an experience that integrates user agency with the engaging power of narrative, and have been recognized as a promising tool for providing both pedagogy and entertainment.

In this paper, we analyze the design desiderata for interactive narrative games and the challenge for creating rich characters in them. We argue that a rich model of characters that are well-motivated, socially aware and have a "Theory of Mind" is needed. In addition, a balance needs to be reached between the design of the characters and the design of the events in the story. We discuss the state of the art work on modeling virtual characters. In particular, we present the Thespian framework for interactive narratives. Thespian utilizes autonomous agents to simulate well-motivated and socially aware characters, which can generate their own behaviors when interacting with the user. During the interaction, Thespian's director agent fine tunes the configuration and behavior of the characters based on the author's plot design and their prior interaction with the user. The director agent does so without affecting the consistency of the characters' motivations. Thus, Thespian is able to create both rich characters and manage the development of the story during the interaction.

REFERENCES

[1] R. Aylett, J. Dias, and A. Paiva. An affectively-driven planner for synthetic characters. In *ICAPS*, 2006.

[2] J. Bruner. *Acts of Meaning.* Cambridge, Mass : Harvard University Press., 1990.

[3] M. Cavazza, F. Charles, and S. J. Mead. Agents' interaction in virtual storytelling. In *Proceedings of the International WorkShop on Intelligent Virtual Agents*, pages 156–170, 2001.

[4] M. Cavazza, F. Charles, and S. J. Mead. Emergent situations in interactive storytelling. In *Proceedings of ACM Symposium on Applied Computing (ACM-SAC)*, March 2002.

[5] S. Chatman. *Story and Discourse: Narrative Structure in Fiction and Film.* Cornell University Press, 1980.

[6] L. Egri. *The Art of Dramatic Writing: Its Basis in the Creative Interpretation of Human Motives.* Simon & Schuster, 2004.

[7] N. Frijda. *The Emotions.* Cambridge University Press, 1987.

[8] J. Gratch and S. Marsella. A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306, 2004.

[9] W. L. Johnson, C. Beal, A. Fowles-Winkler, U. Lauper, S. Marsella, S.and Narayanan, D. Papachristou, and H. Vilhjálmsson. Tactical Language Training System: An interim report. In *7th International Conference on Intelligent Tutoring Systems*, pages 336–345, 2004.

[10] M. T. Kelso, P. Weyhrauch, and J. Bates. Dramatic presence. *Presence: Teleoperators and Virtual Environments*, 2:1–15, 1993.

[11] S. Louchart and R. Aylett. The emergent narrative theoretical investigation. In *the 2004 Conference on Narrative and Interactive Learning Environments.*, 2004.

[12] B. Magerko. A proposal for an interactive drama architecture. In *Working Notes of the AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment*, 2002.

[13] S. C. Marsella, W. L. Johnson, and C. Labore. Interactive pedagogical drama for health interventions. In *AIED*, pages 341–348, 2003.

[14] S. C. Marsella, D. V. Pynadath, and S. J. Read. PsychSim: Agent-based modeling of social interactions and influence. In *Proceedings of the International Conference on Cognitive Modeling*, pages 243–248, 2004.

[15] M. Mateas and A. Stern. Integrating plot, character and natural language processing in the interactive drama Façade. In *the International Conference on Technologies for Interactive Digital Storytelling and Entertainment*, 2003.

[16] K. Nelson. *Narratives from the Crib.* Cambridge, Mass:Harvard University Press., 1989.

[17] E. Ochs and L. Capps. *Living narrative. Creating lives in everyday storytelling.* Cambridge, MA, Harvard University Press, 2001.

[18] S. Onega and J. A. G. Landa. *Narratology: An Introduction.* Longman, London and New York, 1996.

[19] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions. Cambridge.* UK: Cambridge University Press, 1998.

[20] A. Paiva, J. Dias, D. Sobral, R. Aylett, P. Sobreperez, S. Woods, and C. Zoll. Caring for agents and agents that care: Building empathic relations with synthetic agents. In *AAMAS*, pages 194–201, 2004.

[21] D. V. Pynadath and S. C. Marsella. Psychsim: Modeling theory of mind with decision-theoretic agents. In *IJCAI*, pages 1181–1186, 2005.

[22] M. O. Riedl and S. Andrew. Believable agents and intelligent scenario direction for social and cultural leadership training. In *the 15th Conference on Behavior Representation in Modeling and Simulation*, Baltimore, Maryland, 2006.

[23] M. O. Riedl, C. J. Saretto, and R. M. Young. Managing interaction between users and agents in a multi-agent storytelling environment. In *AAMAS*, pages 741–748, 2003.

[24] I. Roseman. Cognitive determinants of emotion: A structural theory. *Review of Personality and Social Psychology*, 2:11–36, 1984.

[25] K. Scherer. Appraisal considered as a process of multilevel sequencial checking. In K. Scherer, A. Schorr, and T. Johnstone, editors, *Appraisal Processes in Emotion: Theory, Methods.* Oxford University Press., Oxford, 2001.

[26] M. Si, S. C. Marsella, and D. V. Pynadath. Thespian: An architecture for interactive pedagogical drama. In *AIED*, pages 595–602, 2005.

[27] M. Si, S. C. Marsella, and D. V. Pynadath. Thespian: Using multi-agent fitting to craft interactive drama. In *AAMAS*, pages 21–28, 2005.

[28] M. Si, S. C. Marsella, and D. V. Pynadath. Thespian: Modeling socially normative behavior in a decision-theoretic framework. In *IVA*, pages 369–382, 2006.

[29] M. Si, S. C. Marsella, and D. V. Pynadath. Proactive authoring for interactive drama: An author's assistant. In *IVA*, pages 225–237, 2007.

[30] M. Si, S. C. Marsella, and D. V. Pynadath. Directorial control in a decision-theoretic framework for interactive narrative. In *Proceedings of International Conference on Interactive Digital Storytelling*, pages 221–233, Guimares, Portugal, 2009.

[31] M. Si, S. C. Marsella, and D. V. Pynadath. Modeling appraisal in theory of mind reasoning. *JAAMAS.*, 2009.

[32] R. D. Smallwood and E. J. Sondik. The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, 21:1071–1088, 1973.

[33] C. A. Smith and P. C. Ellsworth. Patterns of appraisal and emotion related to taking an exam. *Personality and Social Psychology*, 52:475–488, 1987.

[34] C. A. Smith and R. S. Lazarus. Emotion and adaptation. In L. A. Pervin, editor, *Handbook of personality:Theory and research.* Guilford, New York, 1990.

[35] W. Swartout, R. Hill, J. Gratch, W. Johnson, C. Kyriakakis, C. LaBore, R. Lindheim, S. C. Marsella, D. Miraglia, B. Moore, J. Morie, J. Rickel, M. Thiúbaux, L. Tuch, R. Whitney, and J. Douglas. Toward the holodeck: Integrating graphics, sound, character and story. In *Agents*, pages 409–416, 2001.

[36] N. Szilas. IDtension: a narrative engine for interactive drama. In *the 1st International Conference on Technologies for Interactive Digital Storytelling and Entertainment*, pages 14–25, March 2003.

[37] E. S. Tan. *Emotion and the Structure of Narrative Film: Film as an Emotion Machine. Tr. by Barbara Fasting.* Lawrence Erlbaum Associates, Mahwah, N.J., 1996.

[38] D. R. Traum, W. Swartout, S. C. Marsella, and J. Gratch. Fight, flight, or negotiate: Believable strategies for conversing under crisis. In *IVA*, 2005.

[39] F. Wang, K. M. Carley, D. Zeng, and W. Mao. Social computing: From social informatics to social intelligence. *IEEE Intelligent Systems.*, 22(2):7983, 2007.

BIOGRAPHY

MEI SI received her Ph.D. from the Computer Science Department at the University of Southern California. Mei's research focuses on Embodied Conversational Agent, Interactive Narrative and Human-Computer Interaction. She received a M.S. in computer science from the University of Arizona and a M.A. in psychology from the University of Cincinnati. Before coming to the United States, she completed her B.S. in Psychology at Peking University, P. R. China.

TECHNOLOGY OVERVIEW

The All-Round Maturity and Application of CG Technology in Digital Entertainment

Zhiyi Zhang

Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
E-mail: zhiyi.zhang@ia.ac.cn

Abstract

This paper provides an overview of the application of computer graphics (CG) and 3D animation technologies in digital entertainment, in particular in film. With the emergence of 3D imaging era, CG technology is bringing and will continue to bring about great innovations to not only film and TV, but also game and digital entertainment industry in general.

Keywords Computer Graphics, 3D Animation, 3D Film, game, Digital Entertainment

CG TECHNOLOGY FOR ARTISTIC CREATION: FILM AS AN EXAMPLE

The international community will make use of customary computer technology for visual design and production in the area known as computer graphics (CG) technology. CG technology was originated in the fields such as construction and mechanical design. It was later used for the special effects and animations in film, and through this popular art, the public gradually came to understand it. Since the 80s and 90s in the 20^{th} century, CG technology has been developing rapidly. Now CG technology has been applied to almost all areas of the visual art activities such as film special effects, 3D animation, graphic design, multimedia technology, and so on. One of the most familiar areas we know is, of course, its use in the modern film industry. By means of CG technology, film production as an example, can be divided into two categories. One category is “real film” which is the synthetic production of real people and objects through digital technology. The other category entirely uses computer technology to generate virtual images in “cartoon” format.

Both “real film” and “cartoon” are the products of modern industrial civilization. They give the social and cultural values of artistic works. Meanwhile, they are also influenced by the commercial interests of producers. Without violating the fundamental values, the producers typically maximize the business value as their objective. Therefore, the goals that popular art, in particular film, pursues are the “new” (continuously innovating the forms and contents to stimulate the senses of the audience) and the “fast” (technology and management innovation to improve production efficiency). CG technology provides the perfect and uniform way to help film producers achieve these two goals.

Pushing the "Novelty" Winning Rule: The Popularity of Landscape Film

Hollywood, the dominant player in commercial movies, has been occupying most of the world market. Since 1980s, CG technology has brought about a profound change in the film industry. Through CG technology, Hollywood movies have elegantly changed from “narrative cinema” to “landscape film”. Creating a digital “visual spectacle” has been Hollywood’s holy grail to occupy and further consolidate its global market share these years.

A simple glance of CG in the development process of Hollywood "large film" finds that CG technology plays a significant role in promoting popular film. In 1968, Kubrick’s “2001: A Space Odyssey” [1] first employed CG technology, special effects, images and music to dominate the production of this film. Spielberg praised, “I first saw the film. I think this is not a movie. It will change the form of the film.”

In 1982, Lucas’s “Star Wars 2” [2] signified that CG technology began to officially enter the film industry. The film’s 60-second stunt created many “first” in the history of CG film. CG technology attracted the audience’s central attention for the first time. In 1991, Cameron’s “Terminator 2” [3] was the first large-scale use of CG technology in movie. It includes more than 40 computer-generated shots, and the free deformation of the T1000 has become the classic image in sci-fi film history.

Since then, from “Forrest Gump” [4], “Titanic” [5], “Gladiator” [6] and “Lord of the Rings” [7] to recent “Avatar” [8], these films continue to create magnificent scenes of the film feast. A variety of film and television special effects created by CG technology have become a key means to ensure movies at the box office. That is why Cameron said, "The production of visual entertainment video and the corresponding technology are undergoing a revolution. This revolution has changed the way we produce films and other visual media programs so profoundly that we can only describe it by the occurrence of a digital Renaissance."

Promoting the "Efficiency" Winning Rule: The Prevalence of 3D Animation

Traditional animation film has been the "small case" of visual arts as opposed to Hollywood's "large film" investment with "huge amount of money". Except for Disney animated film, investment on traditional types of animation film is risky. This is largely due to the factors that the efficiency of the traditional animation production is very low and the production cycle is usually very long. These factors inevitably result in the high investment risk of animation films.

However, with CG technology, especially the use of three-dimensional animation technology has injected a new vitality to the film animation industry. Digital technology has changed the traditional animation creation and production methods, and pushes film animation industry to a new prosperous era. In 1982, Disney's "TRON" was recognized as the "CG film that creates a new era". Numerous CG industry pioneers were influenced by the film and entered the CG field since then. In 1995, Pixar's "Toy Story" [9] as the first true three-dimensional animation film was born. Since then, Disney, Pixar, DreamWorks, Blue Sky Studios and others have continuously given us visual surprises and box office miracles.

CG and animation technologies provide production staff a broader room for creativity. Production staff can set director and photographer more easily goals to achieve better performance based on their own creative intentions, so that television programs and movies can fully reflect the performance of individual creators. Lucas once said, "In the development of digital technology, my main interest lies in its ability to speed up the process of film production, so that I am able to more effectively realize my ideas. I have been working on the improvement of my film production ability, as currently the major film production process is still the same as that in the 19th century. Digital technology can save a lot of time and energy, which allows filmmakers to be more creative (whether in special effects or in post-production aspects) and to work easier."

Now, whether it is from production methods, artistic expression, or from market response in terms of social impact, many 3D animated films have become the mainstream of development in international animation film industry.

THE EMERGENCE OF 3D IMAGING ERA: "AVATAR" AS AN EXAMPLE

3D Movie - the Real Advent of Film "Experience" Era

In addition to artistic exploration and practice, film development as an audio-visual art, relies critically on the advances of technology. From the maturity of photography that gave birth to movie, sound film technology laid the foundation of audio and visual film. Color film technology changed the way we perceived the world. Digital film technology has made a qualitative change in images to achieve a visual overflight. These advances are all driven by the audience's need for "true illusion".

After "Avatar" screenings, the most frequent word expressed by Avatar's general audience is "virtual reality". In the beginning, most of the media and expert responses are praises and the positive comments on the real advent of three-dimensional movie times. After a period of time some media began to criticize that "Avatar is not the Chinese film's life-saving straw" and questioned "virtual reality": "Is VR the highest realm of film that we need?"

In fact, virtual reality is certainly not the highest realm of commercial films. The highest realm of commercial films is actually the state of the "endless", that is, the endless pursuit of satisfying the audience's need of "true illusion".

Actually, three-dimensional movie is nothing new. Its history is almost as long as cinema itself. The world's earliest 3D movie appeared in the United States in 1922. The first wave of three-dimensional film occurs during 1953 and 1954, when 38 three-dimensional movies were produced in the United States. However, due to technical reasons, early 3D films were ineffective in general.

With the advances of CG technology, previous limitations on film production and popularization as well as a variety of three-dimensional viewing problems have been overcome by the advanced digital technology. The milestone "Avatar" marked the advent of true three-dimensional film era. It is predicted that "compared to 3D film, 2D film is obsolete just the same as silent film to sound film."

This argument has a grain of truth. In the 2D film era, the film frame to some extent is a continuation of the stage box. We can only focus the theater audience's attention on the front screen through dark effect of the theater. No matter how concentrated the audience are, this viewing mode still has the flavor of watching. The emergence of modern 3D film brings about the sense of "experiencing". After watching "Avatar", a spectator felt: "For so many years, I have seen only two films. One is 'Avatar' and the other is all the other movies. Previously I went to the cinema to watch movies, but this time, I feel like to experience the movie."

This sense of "experiencing" still continues the pursuit of "true illusion", and still conforms to the aesthetic principles of modern film such as the "pleasure principle" and the "spectacle principle". The wonderful CG technology is the extremely untrue approach to show the extremely real production methods, so as to create the "virtual real" "spectacles".

The value of products is to meet demands. Consumers' demands are the ultimate driving force of product development. Producers will continue to innovate, manufacture and guide consumers' demands. Film as a popular cultural product, of course owns the properties of general merchandise. When 3D movie is able to better meet the audience's demands, its times comes.

If the audience may be sentimental, and the film comments may be due to biases, then the capital itself is full of rational, and its judgments are all based on numbers. Here are two messages. One is about producers: DreamWorks announced that after 2009, all its produced animated films would use the digital three-dimensional format. Disney made the same announcement as well.

The other message is about cinemas. At present, foreign developed countries have set off a wave of developing digital 3D cinemas. The United States and European exhibitors have started implementing installation programs of digital 3D systems. Until last summer, the whole United States has only 800 3D screens, while from the second half of last year to this year before Avatar's release, the figure was rising rapidly, and suddenly reached 3500. At present, China already has 800 3D screens, and the number is increasing at an astonishing rate.

3D Movie Technology - the Confluence of "Real Film" and "Cartoon"

The most high-end 3D CG technology has brought about an unprecedented shock effect on 3D movies and animation films. But what it really brings about is not only a breakthrough in visual language, but also a revolutionary development. It fully realizes CG film's "creation/production, technical development and process management", the trinity of modern film's industrial production process. From technology and production perspectives, it achieves the confluence of "real film" and "cartoon".

Avatar's creative thinking can be traced back to as early as 20 years ago, when the technology at that time troubled Cameron and also prohibited producers. It is the advances of CG technology as well as Cameron's perseverance that create today's "Avatar". What Cameron has created this time is not just a brilliant film, but the use of the 3D camera, virtual camera, performance capture system, and his firstly attempted GPU accelerated rendering technology. These CG movie technologies created by his team have opened a new era of 3D and digital special effects.

Avatar's production fully draws on the modern three-dimensional animation process and technology. In addition to the lenses acquired via photographing real scenes, other lenses primarily produced via computer technologies through three stages. Digital modeling creates computer models of virtual characters. In the virtual photographing stage, actors are allowed to wear clothes with sensors. By means of computing technology such as performance capture, the parameters of different parts of bodies were collected and used in the virtual models, so that the models are movable. The third stage is rendering. Characters appear in virtual environments by applying rendering technology.

They specifically developed a virtual camera system, with 120 cameras from all directions to capture the actors' performances. Meanwhile, this corresponds to the completed 3D environments in virtual databases. From the camera side, the photographer can see the CG special effects, virtual characters and the scenes after background combination. Rendering is the most time-consuming part of film post-production. In order to deal with the "Avatar" grand scenes and complex coloring work, several sets of coloring systems were connected together via the fiber optic network. Through the use of high-performance GPU cards, real-time computer clustering to process multiple one-level coloring, two-level coloring, multi-point tracking, fuzzy and handling full 3D stereoscopic video workflow, while performing real-time coloring for images in eyes could be achieved.

FUTURE TECHNOLOGY TRENDS

Prospect on 3D Imaging Technology

Current 3D film gives rise to the feeling of three-dimensional images, mainly by the selection of lights from the screen through 3D glasses. The lights are then presented in the left and right eyes and produce the feeling of 3D imaging. From the technical theory point of view, imaging techniques mainly include red and blue filters, polarizing filter imaging technology and 120Hz technology. All of them need special glasses to see the three-dimensional effects. Such technologies are quite mature, and the ornamental effect is quite good.

With the advances of technology, future naked-eye (i.e., no need to wear glasses) true 3D stereoscopic display technology will gradually mature, and the range of applications is limitless. True 3D stereoscopic display is different from our current 3D display based on a two-dimensional flat panel display. It is a new three-dimensional image display technology. Based on this new 3D display technology, we can directly observe three-dimensional images with physical depth. It has the advantages of full visual, multi-angle, simultaneous multi-person observation and real-time interaction, among others. Its display effect of image depth (i.e., the third dimension) is three-dimensional, just as we look at the real world. These technologies help realize this feat, namely, electronic holography, direction-multiplexed display and volumetric display.

Meanwhile, perceptual technology is getting closer and closer to us and will affect our lives in all aspects. Perceptual technology realizes the convenience, accuracy and functional diversity of human-computer interaction. For example, Microsoft is using perceptual technology to develop its next generation system interfaces, including a transparent display device that can perceive hand movement and even eye gaze. Users can operate the device without touching it by hand.

Impact on Film Development

The advancement of three-dimensional imaging and interactive technologies will make significant contributions to film development in the future.

Firstly, new technology creates new cinematic time and space, in which the audience are no longer spectators, but insiders. Full-frame film may cause changes in the concepts of film production and appreciation. For example, currently the concept of appreciation time is a physical one (e.g., 90 or 120 minutes), while it would transform itself into a psychological concept in watching a full-frame film. It is simple to understand this: just as a person who is scared of heights takes a roller coaster ride, he or she would psychologically feel it for a longer time when physically it just takes one minute.

Secondly, new technology produces new ways of media dissemination. Because full-frame film belongs to the spatial category, it doesn't need screens as media carriers. Besides, the feeling of individualized show is very different from that in the traditional watching mode. Traditional acceptance theory in film is built on the basis of traditional cinema watching mode. One big screen, dark scenes and a crowd of spectators are the typical cinema watching mode in film aesthetic studies. However, in watching full-frame film, the space is not limited to the cinema. Crowd watching style is changed to individual/private watching, and this will consequently change the psychology of watching.

Thirdly, new technology makes interactive film possible. In many sci-fi movies, characters can communicate with virtual images without using a telephone. These kind of sci-fi scenes can be designed with interactive technology. The new relation of space and time will certainly change the process of traditional film making. However, it is hard to predict its concrete forms now.

SUMMARY

As far as technology is concerned, lenticular image is mature. Even holographic image and perceptual technology have been developed. However, due to cost and marketing issues, it is not easy to apply them to the actual film making process. The trends of the film industry will depend more on capital than on technology and arts. If there is still a benefit in the traditional film industry, the pace of applying advanced technology in film making will not be very fast. On the other hand, if there is sufficient capital to support new technologies, technology realization will be applied quickly.

No matter when it will come true, as film producers, it is important to be prepared for its coming. So we can stay calm when "Avatar" came. Compared with "Avatar", full-frame film will bring a bigger visual impact in the future. Although many researchers insist that digital technology will change the style of traditional film, and shake the classical film theory by Eisenstein and Andre Bazin, we think that currently, most of the application of digital technology in film is limited to the tool level. So far as in "Avatar", digital technology doesn't change the film's ontology. After all, "Avatar" is shown in cinema, so it is in actual fact a traditional film in the way of appreciation as well as a basis movie in the original sense of the word.

As for the future 3D full-frame movies and interactive movies, they will change film watching style and interactive style. Therefore, they would break the morpha, language and grammar of traditional film. Perhaps this will lead to the creation of a new art language, or even lead to the adjustment of whole film industry, and a new art offspring will be born.

Take a general view of CG technology and reconsider its development in Chinese digital entertainment, we find that CG technology is largely ignored. As film producers, we have a long way to go. As the 'Love crystallization' of arts and technology, CG technology is bringing and will continue to bring about great innovations to not only film and TV, but also game and digital entertainment industry in general.

REFERENCES

1. http://en.wikipedia.org/wiki/2001:_(film). 2001: A Space Odyssey.

2. http://en.wikipedia.org/wiki/Star_Wars. Star Wars.
3. http://en.wikipedia.org/wiki/Terminator_2:_Judgment_Day. Terminator 2: Judgment Day.
4. http://en.wikipedia.org/wiki/Forrest_Gump. Forrest Gump.
5. http://en.wikipedia.org/wiki/Titanic_(1997_film). Titanic.
6. http://en.wikipedia.org/wiki/Gladiator_(2000_film). Gladiator.
7. http://en.wikipedia.org/wiki/The_Lord_of_the_Rings_film_trilogy. Lord of the Rings.
8. http://www.avatarmovie.com/. Avatar.
9. http://en.wikipedia.org/wiki/Toy_Story. Toy Story.

GAMER BEHAVIOUR AND CLASSIFICATION

EXPLORATIONS IN PLAYER MOTIVATIONS : GAME MODS

Barbaros Bostan	Ugur Kaplancali
National University of Singapore	Yeditepe University
Games Lab, 21 Heng Mui Keng Terrace	Kayisdagi Cad. 26 Agustos Yerlesimi
Singapore, 119613	Istanbul / Turkey, 34755
E-mail: idmbb@nus.edu.sg	E-mail: ugur.kaplancali@yeditepe.edu.tr

Abstract

This article aims to analyze player motivations derived from the needs framework of Murray (1938) in relation to user modifications (mods) to an existing commercial computer game. Although the restrictions imposed by the game mechanics significantly reduce the number of player needs satisfied by a game and trap the player within the common motivational cycle of *Achievement*, *Aggression*, *Harmavoidance* and *Acquisition* (Bostan and Kaplancali, 2009), this study shows that the game mods created by users attempt to compensate for this by satisfying the needs of *Sentience*, *Exhibition*, *Recognition*, *Sex*, *Play* and *Affiliation*.

While attempting to find the current trends in user created content for role-playing games (RPGs), this article also discusses the implications of game modding in identifying the missing features of an entertainment experience and of investigating the player motivations. This discussion is framed in terms of the user-environment relations of a recently released popular computer role-playing game (RPG). The gaming experience provided by this genre is analogous to real life and thus has the potential to satisfy a broader range of player needs. The hypothesized mod-need relations are also reviewed with a case study.

Keywords: Player Motivations, Player Psychology, Human Factors, Gameplay Experience, Gaming Mods

PSYCHOLOGICAL NEEDS OF GAME PLAYERS

Game playing is a goal-directed behavior where the needs, motives and goals of a player interact with the opportunities and incentives of the gaming environment and define the concept of "play". Assuming that goal-directed behavior of players is triggered by the interaction between these personal and environmental factors, this article aims to analyze gaming motivations derived from the basic human needs. The psychological needs investigated in this study are based on the psychogenic needs defined by an extensive research of Murray (1938) and consist of six categories. Twenty seven needs of this framework have already been analyzed by Bostan (2009) in relation to the gaming situations of a RPG. In an attempt to take this study one step further and to identify the common interaction patterns between these individual needs, the same motivational framework was applied to another RPG and the individual needs were analyzed by defining the driving game mechanics behind them (Bostan and Kaplancali, 2009). And this study analyzes user-created content (mods) of a popular RPG within the same motivational framework in terms of the needs they satisfy. The RPG selected for this study is *Fallout 3* which was developed and released by Bethesda Softworks in October 2008.

Table 1: Psychogenic Needs of Murray (1938)

Materialistic Needs *nAcq:* Acquisition *nCons:* Construction *nOrd:* Order *nRet:* Retention	**Power Needs** *nAgg:* Aggression *nBlam:* Blamavoidance *nCnt:* Counteraction *nDfd:* Defendance *nDef:* Deference *nDom:* Dominance
Affiliation Needs *nAba:* Abasement *nAff:* Affiliation *nNur:* Nurturance *nRej:* Rejection *nSuc:* Succorance	**Achievement Needs** *nAch:* Achievement *nAuto:* Autonomy *nHarm:* Harmavoidance *nInf:*Infavoidance *nRec:* Recognition *nExh:* Exhibition
Information Needs *nCog:* Cognizance *nExp:* Exposition *nUnd:* Understanding	**Sensual Needs** *nPlay:* Play *nSen:* Sentience *nSex:* Sex

SATISFYING NEEDS THROUGH GAME MODDING

Game modding refers to the activity of making modifications to an existing commercial computer game's aesthetics, experience and structure. This opportunity comes with computer games that are shipped with tools or scripting languages, which allow users to modify the existing virtual world or create new ones. These customized game experiences are usually shared on the Internet and discussed on game forums. Game modding acts as an important source of innovation in the digital games industry which is used as a recruiting pool; it also strengthens the brand name, adds to the shelf-life of the original product, increases customer loyalty and reduces game developers R&D efforts and marketing costs (Kücklich, 2005). From a new media perspective, modding is a perfect example of migration from participatory culture to participatory design (Sotamaa, 2003). These user-made modifications have recently attracted the attention of several researchers. For example El-Nasr and Smith (2006) discusses the use of game modding as a pedagogical activity; Fanning (2006) emphasizes the educational value of game mods; Nieborg (2005) analyzes the current trends in co-created content for First Person Shooter (FPS) games. Mods come in different sizes and complexities, but those that modify a broader range of game elements are named total conversions or overhauls.

From a motivational perspective, game modding is a convenient way of satisfying user needs that are not fulfilled by the game itself. If a human need is a condition marked by the lack of something, game mods are also marked by the lack of something or missing feature in a computer game. The modifications in the game mechanics allow the user to break the constraints imposed by the game, thus providing an opportunity to satisfy his/her broader range of needs. This section will focus on game mods created for *Fallout 3* game. The website selected for the analysis is *Fallout 3 Nexus* [1], which is one of the biggest modding communities for the game, hosting more than 7500 mods. Sixteen categories of mods are chosen from the website, because three categories (saved games/characters, official mods, videos and trailers) are not user-made modifications to the gaming environment.

Before analyzing the game mods of *Fallout 3* from a needs perspective, it is important to note that the unique nature of *Achievement* requires special attention. Since nAch is the dominant psychogenic need that fuses readily and naturally with every other need (Murray, 1938), it will not be specified in the mod-need relations unless it is the only dominant need. The analysis given below also introduces new variables, some of which are not in-game motivational variables. As the variables of the framework analyzed in this study are assumed to be the motivations of players inside the virtual world, the real world motivations of players are represented with the prefix 'r'. For example, an online RPG player's in-game motivation could be to achieve maximum levels in minimum time so that he/she can sell his/her character in auction websites on the Internet and acquire money. The player's in-game motivation is *Achievement*, represented by nAch; but, the player's motivation in the real world is *Acquisition*, which can be represented by rAcq. As the validity of this discrimination is open to question, this issue will be discussed further while dealing with the cases of rExh (the player's desire to attract the attention of people in the real world) and rRec (the player's desire to excite praise and commendation, and to seek distinction). For each category, the 10 most downloaded mods are tested and analyzed with respect to the needs it satisfies. Table 2 below shows 160 mods of 16 categories listed in descending order, considering the total number of downloads (See Appendix for details).

Table 2: Gaming mods and motivational relations

Category: Animation	
ID	**Needs Satisfied**
M1	nSen
M2	nSen, AND/OR rExh, rRec
M3	nSen, AND/OR rExh, rRec
M4	nSen, varies
M5	nSen, varies
M6	nSen, AND/OR rExh, rRec
M7	nSen, AND/OR rExh, rRec
M8	nSen
M9	nSen, AND/OR rExh, rRec
M10	nSen
Category: Armors	
ID	**Needs Satisfied**
M11	nSen, nSex, AND/OR rExh, rRec
M12	nSen, nAgg, nHarm, AND/OR rExh, rRec
M13	nSen, nHarm, AND/OR rExh, rRec
M14	nSen, nSex, AND/OR rExh, rRec
M15	Varies
M16	nSen, nSex, AND/OR rExh, rRec
M17	nSen, nSex, AND/OR rExh, rRec
M18	nSen, nSex, AND/OR rExh, rRec
M19	nHarm
M20	nSen, nAgg, nHarm, AND/OR rExh, rRec
Category: Buildings	
ID	**Needs Satisfied**
M21	nSen, nRet, nOrd, nAcq, nHarm, nCons
M22	nAcq, varies
M23	nSen, nRet, nOrd, nAcq, nHarm, nCons
M24	nSen, nRet, nOrd, nAcq, nHarm, nCons
M25	nSen, nRet, nOrd, nAcq, nHarm, nCons
M26	nSen, nRet, nOrd, nAcq, nHarm, nCons
M27	nSen, nRet, nOrd, nAcq, nHarm, nCons
M28	nSen, nRet, nOrd, nAcq, nHarm, nCons
M29	nSen, nRet, nOrd, nAcq, nHarm, nCons
M30	nSen, nRet, nOrd, nAcq, nHarm, nCons
Category: Cheats and God Items	
ID	**Needs Satisfied**
M31	nRet
M32	nAcq, nRet
M33	nAch, varies
M34	nAch, varies
M35	nAch, varies
M36	nAcq, nRet
M37	nAcq, nRet
M38	nAch, varies
M39	nAch, varies
M40	nAcq AND/OR nHarm
Category: Clothing	
ID	**Needs Satisfied**
M41	nSen, nSex, AND/OR rExh, rRec
M42	nSen, nSex, AND/OR rExh, rRec
M43	nSen, nAcq, nRet
M44	nSen, nSex, AND/OR rExh, rRec
M45	nSen, nSex, AND/OR rExh, rRec
M46	nSen, AND/OR rExh, rRec
M47	nSen, AND/OR rExh, rRec
M48	nSen, AND/OR rExh, rRec
M49	nSen, nSex, AND/OR rExh, rRec
M50	nSen, nSex, AND/OR rExh, rRec
Category: Companions	
ID	**Needs Satisfied**
M51	nAff, nAgg
M52	nAff, nSen
M53	nAff, nSen
M54	nAff, nAgg
M55	nAff, nSen, nSex, nAgg

M56	nAff
M57	nAff, nAgg
M58	nAff, nAgg
M59	nAff, nAgg
M60	nAff, nSen
Category: Gameplay Effects and Changes	
ID	**Needs Satisfied**
M61	nAch, varies
M62	nPlay
M63	nAgg, nHarm
M64	nAgg, nAcq, nSen
M65	varies
M66	nAgg, nHarm
M67	nAff, nAgg
M68	nSen
M69	nRet
M70	nAgg
Category: Guilds / Factions	
ID	**Needs Satisfied**
M71	nAcq
M72	nAgg, nAcq
M73	nAcq
M74	nPlay
M75	nAff, nAcq
M76	nAcq
M77	nHarm
M78	nAff, nAcq
M79	nSen, nRet, nOrd, nAcq, nHarm, nCons
M80	nPlay
Category: Hair and Face Models	
ID	**Needs Satisfied**
M81	nSen, AND/OR rExh, rRec
M82	nSen, AND/OR rExh, rRec
M83	nSen, AND/OR rExh, rRec
M84	nSen, AND/OR rExh, rRec
M85	nSen, AND/OR rExh, rRec OR nAgg
M86	nSen, AND/OR rExh, rRec
M87	nSen, AND/OR rExh, rRec
M88	nSen, AND/OR rExh, rRec
M89	nSen, AND/OR rExh, rRec
M90	nSen, AND/OR rExh, rRec
Category: Miscallenous	
ID	**Needs Satisfied**
M91	varies
M92	nAch, varies
M93	nAch, varies
M94	nAch, varies
M95	nPlay
M96	nAch, varies
M97	nAff
M98	nAch, varies
M99	nPlay
M100	varies
Category: Models and Textures	
ID	**Needs Satisfied**
M101	nSen, nSex, AND/OR rExh, rRec
M102	nSen, nSex, AND/OR rExh, rRec
M103	nSen

M104	nSen, AND/OR rExh, rRec
M105	nSen, nAcq
M106	nSen, AND/OR rExh, rRec
M107	nSen, nSex, AND/OR rExh, rRec
M108	nSen
M109	nSen, nSex, AND/OR rExh, rRec
M110	nSen
Category: New Lands	
ID	**Needs Satisfied**
M111	nSen, nAcq, varies
M112	nSen, nRet, nOrd, nAcq, nHarm, nCons
M113	nAgg, nAcq
M114	nAgg, nAcq
M115	nSen, nRet, nOrd, nAcq, nHarm, nAgg, nCons
M116	nSen, nAcq
M117	nSen, nRet, nOrd, nAcq, nHarm, nCons
M118	nSen
M119	nSen, nRet, nOrd, nAcq, nHarm, nCons
M120	nSen, nRet, nOrd, nAcq, nHarm, nCons
Category: NPCs	
ID	**Needs Satisfied**
M121	nAgg
M122	nAgg
M123	nSen, nAff
M124	nSen
M125	nAch, varies
M126	nAgg
M127	nAgg, nAcq
M128	nSen
M129	nSen, nAff
M130	nAff
Category: Quests and Adventures	
ID	**Needs Satisfied**
M131	nDom, nSex, nAcq
M132	nSex
M133	nDom, nNur OR nSex,
M134	nAgg, nAcq
M135	nAcq, nAff, nSen
M136	nSex
M137	nAcq
M138	nAcq, nDom
M139	nAcq
M140	nAgg, nAcq
Category: Sounds and Music	
ID	**Needs Satisfied**
M141	nSen
M142	nSen
M143	nSen
M144	nSen
M145	nSen
M146	nSen
M147	nSen
M148	nSen
M149	nSen
M150	nSen
Category: Weapons	
ID	**Needs Satisfied**
M151	nAgg, nSen

M152	nAgg, nSen
M153	nRet
M154	nAgg, nSen
M155	nAgg, nSen
M156	nAgg, nSen
M157	nAgg, nSen
M158	nAgg, nSen
M159	nAgg, nSen
M160	nAgg, nSen

Most of the mods under the category of "animation" are related to character poses (M1, M2, M3, M6, M7, M8, M9). As physical appearance of the player does not impact on the NPCs (non-player characters) and the gaming environment is not a multi-player one, the motivation behind giving fancy or attractive poses is not to excite, stir, shock or thrill virtual characters. People may use these poses (as well as some of the mods related to "clothing", "armor", "hair and face models", and "models and textures") to enhance their attractiveness and customize their physical appearance. Later, they may take screenshots from the game and share the pictures of their characters on the Internet to attract the attention of others (rExh), and to seek distinction and elicit praise (rRec). There are even clothing contests on gaming websites where players post the screenshots of their characters for recognition and exhibition. It should be noted that if NPCs are responsive to the physical appearance or the attractiveness of the player, these mods should also have some in-game motivation behind them (nRec, nExh). Players are also attracted by pleasurable sights (color, light, form, pose, movement, a beautiful face, clothes, etc.) and they seek variety in these sensuous impressions (nSen). To delight in the beauty of one's own body and to enjoy sensuous imagery is a special form of sentience called intraSentience (Murray, 1938, p. 169). The rExh-rRec-nSen pattern[2] can also be seen in "armors", "clothing" and "models & textures". Two mods of this category (M4, M5) modify the mechanics behind the exploration of the virtual environment.

Figure 1: Player character posing in front of the statue of Lincoln at Lincoln Memorial[3].

Armors (M12, M20) can boost combat skills (nAgg) and increase damage resistance (nHarm). They can help the player avoid harm by disguise or stealth (M13, M19), but half of the mods under this category (M11, M14, M16, M17, M18) introduce nSex into the motivational analysis. What is experienced is the common fusion of eroticism with exhibition, called exhibitionism (Murray, 1938, p.168). These sensuous (nSen) and sexually revealing armors are usually used with body textures that permit replacement of female models in default underwear with nude ones. These modifications are in-line with the way female characters are commonly portrayed in computer games. Researchers analyzing gender stereotyping in video game characters observe that female characters are usually hypersexualized, depicted as sex objects, dressed in sexually revealing clothing, featured in unrealistic, curvaceous or voluptuous body shapes with special emphasis on their breasts and displayed in camera angles that reveal their clothing or shapely bodies (Provenzo, 1991; Gailey, 1993; Dietz, 1998; Inness, 1999; Children Now, 2001; Thompson, 2002; Brand, Knight & Majewski, 2003; Downs & Smith, 2005; Jansz & Martis, 2007). The same trend[4] (nSen, nSex) is also evident in the mods of "clothing" and "models and textures".

Figure 2: Player character with Army Dress in front of the Capitol Building[5].

Nine mods, under the category of "buildings", add new structures to the gaming world that are designed to accommodate the player. These buildings are decorated with storage containers (nRet), items such as weapons, ammunition, etc. (nAcq), workbenches for making items (nCons), jukeboxes and radios for listening to music and radio stations (nSen), laboratories for brewing drugs and clearing addictions (nHarm), infirmaries for resting and reducing radiation levels (nHarm), and robot butlers for changing player's hair style or appearance (nSen). Although the game provides six different themes (Love Machine, Science, Vault Raider, Wasteland Explorer, and Pre-War) for the house owned by the player[6], the visual appearance of the buildings is usually customized by each modder. These visual customizations and the addition of more containers for hoarding items are two dominant features. The remaining mod (M22) under this category is a shop that sells armor to the player. Most of the mods, under the category of “cheats and god items”, include lists of cheats available in *Fallout 3*. The cheats can be used to add items, skill points, perks or experience points to the player, which satisfy the player’s needs of achievement and acquisition. Some of them are used for preserving items (nRet), and one for even eliminating the need for item repairs (M31). The motivational aspects (rExh, rRec, nSen, nSex) of the mods, under the category of “clothing”, are similar to the mods of “poses” and “armors”. The only exception (M43) is the one that allows the player to carry more items and hoard them.

Companions in *Fallout 3* are followers or joinable NPCs that come with their own weapons and armor. Their primary purpose is to aid the player in combat and they are hostile towards all enemies. It is questionable whether they satisfy the need for affiliation, but they definitely satisfy the need for aggression, because they are so programmed[7]. It is intriguing to see three mods (M52, M53, M60) that are dedicated to a single NPC. This female character, called Bittercup, is a citizen of Big Town with a “goth” or “punk” teenage personality wearing dark clothes. A little distinction in style and attitude of a character has drawn the attention of modders who try to customize this NPC’s physical appearance (nSen). Mods under the category of “gameplay effects and changes” serve different purposes. Two of them are overhauls (M61, M65), and the others add more features (auto-fire mechanism, extended magazines/clips, etc.) to weapons (M63), modify the spawn frequency of creatures and provide diversity in creature types (M64, M70), add more critical affects to the game (M66), add robot companions (M67), change the visual effects of weather (M68), and modify item degradation (M69). In *Fallout 3*, when the final quest is finished, the players cannot continue playing the game. Therefore, one mod (M62) is specially designed to allow players to continue their adventures (nPlay) in the post-apocalyptic world of Fallout.

Figure 3: Bittercup NPC companion of “Better Bittercup Restyled” mod[8].

Guilds and factions are important parts of the social system of RPGs. They form a group of NPCs that share their resources and help each other. Player may also be given quests for advancement in guild ranks. Mods, under the category of “guilds and factions”, add variety to NPC’s equipment (nAcq) which can be looted by the player (M71, M76), add more rewards for capturing/killing NPCs (M72, M73), and add new factions to the game world (M75, M78). One mod (M77) enhances disguise (actually a “gameplay effect and change”) and another (M79) adds a new home for the player (actually a “building”). Two mods create an alternative gameplaying experience (M74, M80), allowing the player to play a paladin or a ghoul (nPlay). "Hair and face models" provide new faces or hair styles to the players, enabling them to customize and enhance their appearance and attractiveness. "Miscallenous" mods serve different needs. The most downloaded one (M91) is the flash map showing all the locations of the game, which can be used for various purposes. Besides cheat codes, patches and fixes (M95, M96, M99, M100), four of them (M92, M93, M94, M98) remove the level cap (level or skill limit set by designers) and allow the player to progress further in the game (nAch). One interesting mod (M97) allows the players to listen to Galaxy News Radio Station even after 20th level. Run by the DJ, named Three Dog, possibly referring to the 1960's radio personality "Wolfman Jack"[9], the radio station not only plays music but also reports on the player exploits, thus creating a strong social bond with the player.

"Models and textures" change the appearance of the terrains (M103, M108), weapons (M110), and the face of the player character (M106). One mod (M104) enhances female character's attractiveness by makeup effects and one (M105) - actually an item overhaul - adds new items with new textures. And, four of them (M101, M102, M107, M109) are replacements for nude textures and enhancements in female bodies, such as voluptuous breasts, etc. "New Lands" add new places such as an island (M111), a canyon (M113), a single building (M112, M114, M115, M117, M119) or a cavern (M120). Two of them (M116, M118) are modifications of existing towns. The visual appearance of these towns is customized by modders, with additions of items to loot (nAcq), containers to use (nRet) or creatures to kill (nAgg). Mods, under the category of NPCs, add new enemies to kill (M121, M122, M126, M127), characters/monsters with visual enhancements (M123, M124, M128) one of which is also a companion (M129), a trainer to help the player to level up (M125) and two butler robots for the player's house (M130). Children in *Fallout 3* can not be killed. Whether “Killable Children” mod is designed for the sadistic purpose of killing children or simply for lifting another restriction of the game is not clear. And, a better Bittercup is again introduced in this category of mods. "Quests and adventures" introduce new enemies to eliminate (M134, M140) or new items to acquire (M135, M137, M138, M139). Four mods of this category are related to the need for sex, allowing the player to seduce men or women (M132, M136), to persuade female NPCs into becoming prostitutes to reap the benefits of sexual favors or weekly payments (M131), or to blackmail or save a prostitute of the world (M133). Mods of "sounds and music" add new sound effects to the game world or new songs for the Galaxy News Radio, which are sensual (auditory) enhancements to the gaming experience. And mods of "weapons" category add new weapons with new textures that can inflict more damage to enemies. The only exception is the one (M153) that changes the basic principle of item repairs.

CONCLUSION

Although the high degree of technical competence demanded places modding practices out of reach of many gamers, game modding creates wider cultures, communities and rich contexts for criticism, review, and play, and modders are positively encouraged with tools, support and means of distribution by official game developers (Newman, 2008). Needless to say, the sole purpose of creating game mods is not to merely satisfy some user needs. Game mods can also be designed to meet certain virtual environment design requirements, such as interactivity, sociability, veridicality, responsiveness, autonomy, etc., which potentially increase the sense of presence and quality of immersion. And it is also obvious that the relative importance of the needs may change from one game/genre to another, but the variables of this taxonomy could assist the analysis of gaming experiences within a motivational framework. Given below is the pie chart showing the needs satisfied by the gaming mods analyzed in this study.

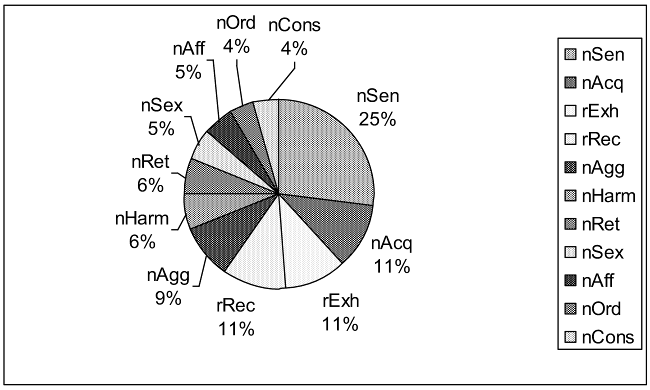


Figure 4: Percentage of needs satisfied by the 160 mods.

The psychological impact of game mods is not limited to the satisfaction of some user needs. These user made modifications to the virtual environment also have an effect on certain components of goal-directed behavior or flow experiences, such as increasing the valence of the incentives, changing the difficulty and specificity of goals, and increasing user commitment and concentration. As game modding requires basic understanding of 3D objects and scripting techniques, it still remains the privilege of an elite group of players. It is not possible to expect these mods to satisfy all user needs or to cover all the flaws of a game perfectly, but when analyzed from a player perspective, they might give some idea about player's choices or preferences in a virtual world. Even if the modders have other goals in mind when designing the mods, it is still possible to investigate what user needs they can satisfy. Given below is the top 10 mods analyzed in this study.

Table 3: Top 10 mods of *Fallout 3*[10].

TOP 10 MODS		
ID	**Downloads**	**Needs Satisfied**
M101	128634	+nSen, +nSex, AND/OR +rExh, +rRec
M61	83552	+nAch, varies
M141	82897	+nSen
M81	81393	+nSen, AND/OR +rExh, +rRec
M131	75797	+nDom, +nSex, +nAcq
M91	62723	varies
M62	59197	+nPlay
M92	54029	+nAch
M102	48636	+nSen, +nSex, AND/OR +rExh, +rRec
M103	47908	+nSen

To test the hypothesized associations between gaming mods and user needs, undergraduate students from a major private university in Turkey who had prior gaming experience in Fallout 3 and its gaming mods were called to participate in a case study. Seven students with the desired gaming background responded to the call. A 10 minute presentation was given to the students to describe the needs framework of Murray. Then, each student was given a score sheet to evaluate the relationship between 27 psychogenic needs and ten randomly selected mods. Each mod was portrayed with a short description and a screenshot. The scoring is between 0 and 10, where a 0 means that the mod does not satisfy the need in question and a 10 means that the mod perfectly satisfies it. The seven students spent 1325 points to describe the mod-need relationships. 909 of these fall in the hypothesized need relationships of this study which represents a 68.6% recognition rate which backs up the applicability of Murray's framework to computer gaming. Although the restrictions imposed by game mechanics significantly reduce the number of player needs satisfied by a computer game, this study shows that gaming mods satisfy a broader range of needs and provide valuable information on player preferences. Understanding these preferences should facilitate the identification of independent player profiles or playstyles and if the hypothesized need-mod relations can be validated by a larger scale empirical study, the framework defined in this study should facilitate the discrimination of a satisfying gaming experience from an unsatisfying one.

REFERENCES

Brand, J.E.; Knight, S.J.; and Majewski, J. 2003. The diverse worlds of computer games: a content analysis of spaces populations, styles and narratives. In *Digital Games Research Conference 2003*, University of Utrecht, The Netherlands.

Bostan, B. 2009. "Player Motivations: A Psychological Perspective." *ACM Computers in Entertainment,* 7, No.2 (June).

Bostan, B. and Kaplancali, U. 2009. "Explorations in Player Motivations: Game Mechanics". In *Proceedings of GAMEON 2009, Duesseldorf, Germany.*

Children Now. 2001. "Fair play? Violence, gender and race in video games". Available on-line at

http://publications.childrennow.org(accessed 19 January 2009).
Dietz, T.L. 1998. "An examination of violence and gender role portrayals in video games: Implications for gender socialization and aggressive behavior". *Sex Roles, 38*, 425–442.
Downs, E. and Smith, S. 2005. Keeping abreast of hypersexuality: A video game character content analysis. Paper presented at the *55th International Communication Association*, New York.
El-Nasr, M.S. and Smith, B. 2006. "Learning Through Game Modding". *Computers in Entertainment (CIE), 4*(1), Article 3B.
Fanning, E. 2006. "Game Mods: Customizable Learning in a K16 Setting". *International Journal of Information and Communications Technology Education (IJICTE),2* (4).
Gailey, B. 1993. "Mediated messages: Gender, class, and cosmos in home video games". *Journal of Popular Culture, 27*, 81–97.
Inness, S.A. 1999. *Tough Girls: Women Warriors and Wonder Women in Popular Culture*. University of Pennsylvania Press, Philadelphia, PA.
Jansz, J. and Martis, R.G. 2007. The Lara phenomenon: Powerful female characters in video games. *Sex Roles, 56*, p. 141-148.
Kücklich, J. 2005. "Precarious Playbour: Modders and the Digital Games Industry". *Fibreculture Journal, 5*. http://www.journal.fibreculture.org/issue5/kucklich.html (accessed January 23, 2009).
Murray, H.A. 1938. *Explorations in personality*. New York: Oxford University Press.
Newman, J. 2008. *Playing with Videogames*. London: Routledge.
Nieborg, D.B. 2005. "Am I Mod or Not? - an Analysis of First Person Shooter Modification Culture". *Paper presented at Creative Gamers Seminar - Exploring Participatory Culture in Gaming*. Hypermedia Laboratory, University of Tampere.
Provenzo, E.F. 1991. *Video Kids: Making sense of Nintendo*. Cambridge, MA: Harvard.
Sotamaa, O. 2003. "Computer game modding, intermediality, and participatory culture". Presented at *New Media? New Theories? New Methods? Innovating Media and Communication Research*. http://old.imv.au.dk/eng/academic/pdf_files/Sotamaa.pdf (accessed December 16, 2008).
Thompson, C. 2002. Violence and the political life of video games. In *Game on: The history and culture of videogames*, 2002, L. King (Ed.). London, UK: Barbican.22–32.

APPENDIX

Category: Animation		
Mod Name	**ID**	**Download**
Trynfkwtmes Exotic Dancers	M1	17586
Umpa Animation	M2	15935
Some Poses	M3	9896
Jump	M4	6099
Drivable Motorcycle	M5	4482
PC Idles	M6	4204
The Groovatron	M7	3728
Alternative Weapon Attack Anim.	M8	2849
Custom Reload Animations	M9	1196
Creature Poses	M10	776
Category: Armors		
Mod Name	**ID**	**Download**
Trynfkwtmes Nude Armors	M11	32527
Better Power Armor	M12	13994
Coyote Reflex Power Armor	M13	13337
Malos Nude Fixes	M14	10987
Missing Unique Armor and Clothing	M15	9994
Real Latex	M16	9497
Naouak Armors and Clothes	M17	9230
Malos Armors and Dresses	M18	8713
Armor Disguises	M19	6944
Crusaders Armor - Power Armor	M20	6231
Category: Buildings		
Mod Name	**ID**	**Download**
Megaton House Overhaul	M21	14366
Trynfkwtmes Armor Shop	M22	8813
RR Companions Vault	M23	8326
Vault 89_player home_V1-26	M24	5691
A Follower Friendly Personal Vault	M25	5017
Dtoms Rivet City Apartment	M26	4909
KORs Vault	M27	4504
My Megaton Basement	M28	4116
My Megaton Basement	M29	3429
Fort Freeway	M30	3346
Category: Cheats and God Items		
Mod Name	**ID**	**Download**
No Repair Needed	M31	15052
Mega Weight	M32	13822
Fallout 3 Cheat Guide	M33	10505
Levelers Truck Trailer	M34	6062
Ultimate Code Guide	M35	5798
Add All Uniques	M36	5588
Choose what you need	M37	4963
Bobblehead Delivery	M38	4722
Cheat List Sorted	M39	3488
Lockpick cheat	M40	3323
Category: Clothing		
Mod Name	**ID**	**Download**
Female Nude Painspike Armor	M41	18054
Just Undies by Azar	M42	12321
Wearable Blackwolf Backpack	M43	12308
Gypsy Outfits by Azar	M44	11744
PMAY Sexy Sleepwear	M45	8376
WW2 Uniforms	M46	7509
Merc Duster	M47	5949
Zumbs Dark Sunglasses	M48	4939
Nippy Clothing	M49	4813
Black Outfit Beta	M50	4811
Category: Companions		
Mod Name	**ID**	**Download**
Enclave Commander	M51	35740
Bittercup companion	M52	23703
Bittercup companion restyled	M53	19879
0h Followers Hire	M54	17696
Jessi Companion	M55	14674
Sharing and Caring Companions	M56	13701
Buildable Bots	M57	13359
CRI Squad	M58	11503
Essential Followers and Caravans	M59	10167
Bittercup Companion Restyled	M60	9434
Category: Gameplay Effects and Changes		
Mod Name	**ID**	**Download**
Ultimate Perk Pack	M61	83552

Free Play after MQ	M62	59197
Weapon Mod Kits	M63	46485
Marts Mutant Mod	M64	35326
Fallout 3 Balance Overhaul	M65	31852
Amplified Crippled Effects	M66	27258
RobCo Certified	M67	24707
Fallout 3 Weather Overhaul	M68	23471
Slower Item Degradation	M69	21381
MMMF3 Increased Spawns	M70	19859
Category: Guilds / Factions		
Mod Name	**ID**	**Download**
Raiders Regulators Talon Expanded	M71	4217
Lawbringer Bounties	M72	1831
More Outcast Rewards	M73	1824
Alternative Beginning - Paladin	M74	1344
US Army	M75	1280
Talon Company Standard Issue	M76	849
Disguiser	M77	741
Regulators Ride Again	M78	619
Naval Command Bunker	M79	534
Goul Playable	M80	434
Category: Hair and Face Models		
Mod Name	**ID**	**Download**
Hair Pack	M81	81393
Josef Greys Kozaburo Hair Style	M82	31483
Spartan VIs Pretty Girls	M83	14241
Dimonized presets for Vanilla races	M84	6834
Enhanced Children Project	M85	6381
Enable All Hairs	M86	5244
Any Hair You Want	M87	5226
Faces of the Law	M88	3411
Frederyck New Female Face	M89	2910
Choose a Character	M90	2659
Category: Miscallenous		
Mod Name	**ID**	**Download**
Fallout 3 Flash Map	M91	62723
Level Cap Increaser	M92	54029
Level Cap Increaser Skill Workaro.	M93	26948
Level 100 Cap	M94	15789
Unofficial Fallout 3 Patch	M95	13796
Console Codes	M96	9667
GNR fix	M97	7040
Level-Up No Lock	M98	6958
Fallout 3 Tweak Guide	M99	6818
Russian 1C Localizations	M100	6038
Category: Models and Textures		
Mod Name	**ID**	**Download**
Brurpo Alluring Body Enhancement	M101	128634
Malos Nude Body Replacer	M102	48636
Terrain pack	M103	47908
Josef Greys Makeup Face Retexture	M104	44219
FOOK - FallOut 3 Overhaul	M105	40723
Female Texture Mod	M106	32465
Naouak Female Body Replacer	M107	30320
Mimezu texture project	M108	27603
Dimonized female races and textures	M109	27222
Hi-Res Weapons	M110	26825
Category: New Lands		

Mod Name	ID	Download
JaySuS Commonwealth Of Modders	M111	16380
Megaton House Expansion Mod	M112	8304
Death Canyon	M113	7056
Free Building	M114	6319
Ghoul Mansion	M115	6239
Big Town Upgraded	M116	6118
New Home Outdoor Bus	M117	3730
Arefu Expanded	M118	3580
Two new wastelandish homes	M119	3019
Megaton caverns	M120	1977
Category: NPCs		
Mod Name	**ID**	**Download**
Killable Children	M121	25804
Boss Monster Addon	M122	5577
Tchos Better Bittercup	M123	3228
Lil Critters	M124	3019
Perk Trainer	M125	2583
Eagle Strikes the Nest	M126	2452
Raiders Redone	M127	2376
NPC Height Randomizer	M128	2353
Maria-A latina sniper	M129	2138
Wadsworth and Godfrey	M130	1801
Category: Quests and Adventures		
Mod Name	**ID**	**Download**
Wasteland Player	M131	75797
Seducing Women	M132	10904
Saving Nova	M133	7362
D6 Wasteland Boss Pack	M134	6525
WReality Haven	M135	6051
Wasteland Gigolo	M136	5274
Quest incl G36 MSG90	M137	4979
101 The Collector	M138	4800
Megaton Contracts	M139	4216
Darkness	M140	3454
Category: Sounds and Music		
Mod Name	**ID**	**Download**
GNR More Where That Came From	M141	82897
Realistic Gunshot Sounds	M142	29987
20 more aged songs	M143	7922
GNR Radio Fixed Extended	M144	7911
Custom Galaxy Radio	M145	7690
IPip Player	M146	7682
Talk Thing with Cruk - Mutant Radio	M147	7519
Improved Sound FX	M148	6405
GNR Extended	M149	5174
Radio Station Fix	M150	4197
Category: Weapons		
Mod Name	**ID**	**Download**
Better Arsenal Desert Eagles V5	M151	33689
Classic Fallout Weapons	M152	20613
Cross-Repair Weapons	M153	20567
Beretta M9 -Release-	M154	16142
Frederyck Tactical Weapons	M155	15147
Pimp my gun Desert Eagle Clean	M156	14713
Better Arsenal FN FAL V5	M157	14655
M1911 UV added	M158	13612

M88 Gauss Rifle	M159	10989
Lightsaber 0_9	M160	10866

[1] URL: (http://www.fallout3nexus.com), downloads as of March 14, 2009.

[2] To display beauty is a fusion of nExh and nSen (Murray, 1938, p. 171)

[3] Special thanks to Yoshikinakota (creator of "Hair Pack" mod) for the hair model, Backsteppo (creator of "Some Poses" mod) for the pose.

[4] The sensations to excite an erotic feeling create a fusion of nSen with nSex (Murray, 1938, p. 169).

[5] Special thanks to Yoshikinakota (creator of "Hair Pack" mod) for the hair model, JosefGrey (creator of "Josef Greys Makeup Face Retexture" mod) for makeup effects, Kalten1979 (creator of "Malos Armors and Dresses" mod) for the Army Dress, Backsteppo (creator of "Some Poses" mod) for the pose.

[6] The player is awarded a Megaton House or a Tenpenny Tower Suite at the end of "The Power of Atom" quest.

[7] To fight together against a common enemy is a fusion of nAgg and nAff (Murray, 1938, p. 175).

[8] Special thanks to Vaungh (creator of "Bittercup Companion Restyled" mod) for the restyled Bittercup.

[9] URL: http://en.wikipedia.org/wiki/Wolfman_Jack

[10] Downloads as of March 14, 2009.

THE CHALLENGE OF BELIEVABILITY IN VIDEO GAMES: DEFINITIONS, AGENTS' MODELS AND IMITATION LEARNING

Fabien Tencé*,**, Cédric Buche*, Pierre De Loor* and Olivier Marc**
* UEB – ENIB – LISyC
** Virtualys
Brest – France
{tence,buche,deloor}@enib.fr

ABSTRACT

In this paper, we address the problem of creating believable agents (virtual characters) in video games. We consider only one meaning of believability, "giving the feeling of being controlled by a player", and outline the problem of its evaluation. We present several models for agents in games which can produce believable behaviours, both from industry and research. For high level of believability, learning and especially imitation learning seems to be the way to go. We make a quick overview of different approaches to make video games' agents learn from players. To conclude we propose a two-step method to develop new models for believable agents. First we must find the criteria for believability for our application and define an evaluation method. Then the model and the learning algorithm can be designed.

KEYWORDS

Adaptive decision making, believability, human-like, evaluation, imitation learning, video games.

INTRODUCTION

Nowadays, more and more consoles and video games are designed to make the player feel like he/she is in the game. To define how well this goal is achieved, two criteria have been defined in academic research: *immersion* and *presence*. According to Slater, immersion is an objective criterion which depends on the hardware and software (Slater et al. 1995). It includes criteria based on virtual sensory information's types, variety, richness, direction and in which extend they override real ones. For example, force feedback and motion sensing controllers, surround sound and high dynamic range rendering can improve the immersion. Presence, also known as telepresence (Steuer 1992), is a more subjective criterion. It is defined as the psychological sense of "being there" in the environment. It is mainly influenced by the content of the video game.

As stated in (Slater et al. 1995), presence partly depends on the match between sensory data and internal representation. This match expresses the fact that we try to use world models to better understand what we perceive and to be able to anticipate (Held and Durlach 1991). This idea is close to what is called *believability* in the arts. Indeed, we can believe in fictional objects, places, characters and story only if they mostly fit in our models. Enhancing believability of video games' content should then enhance the presence.

As there are many ways to enhance believability in video games, we choose to focus on believable virtual characters, also known as believable agents. The reason why we make this choice is because characters have often a major role in the believability of book and movie stories. However, unlike book and movie characters, agents should be able to cope with a wide range of possible situations without anyone to tell them what to do. Instead of defining manually these behavior, it can be interesting for the agents to learn them, reducing the time to design a character. The ability to learn has also the advantage of increasing believability so it should be considered as a must-be feature.

The goal of this paper is to have an overview of the possibilities and constraints for the realization of a believable and learning-capable agent in a video game. We first define what are believable agents in games, how we can evaluate them and the relations between believability and the type of video games. Then, we make a quick overview of how an agent can express believable behaviors and how it can learn them. Finally, we conclude on a two-step protocol to realize believable agents.

BELIEVABILITY

Definition of Believability

The notion of believability is highly complex and subjective. To define and understand this concept, we must look at its meaning in the arts where it is a factor of *suspension of disbelief* (Bates 1992). According to Thomas and Johnston, two core animators of Disney, believable characters' goal is to provide the *illusion of life* (Thomas and Johnston 1981). Reidl's definition is more precise:

"*Character believability refers to the numerous elements that allow a character to achieve the 'illusion of life', including but not limited to personality, emotion, intentionality, and physiology and physiological movement*" (Riedl and Young 2005, page 2). Loyall tries to be more objective saying that such a character "*provides a convincing portrayal of the personality they [the spectators] expect or come to expect*" (Loyall 1997, page 1). This definition is quite close to one factor of the presence, the match between players' world model and sensory data.

If we want to apply the believability definition for video games, things become even more complex. Unlike classic arts, players can be embodied in a game by the mean of virtual bodies and can interact. The believability question is now: does a believable character have to give the illusion of life or have to give the illusion that they are controlled by a player? (Livingstone 2006). There can be very important differences as even if the video game depicts the real world, all is virtual and players know that their acts have no real consequence.

In this paper, we will consider only believable as "giving the illusion of being controlled by a player". At first glance, it can be seen as going against presence as we remind the players that there is a real world. However, not using this approach has some drawbacks too: virtual characters only giving the illusion of life can be classified as "being a piece of program" which may break the illusion permanently. Players may also see problems in the virtual characters' behavior only because they know they are not human-controlled.

Now that we chose a definition for believability, we have to find the different criteria that influence it. Defining those criteria will make improvements and evaluation possible.

Believability Criteria

Criteria to define believability are highly domain-dependant. For example, for embodied conversational agents (ECA), factors like quality of the speech, facial expressions, gestures are important (Ruttkay et al. 2002). For emotional and social agents, reaction to others is the most important part (Reilly 1996) and so on with the numerous domain working on believable agents. However, there is one criteria that can be used for every domains. What have been named Eliza effect (Weizenbaum 1966) is really interesting: as long as the agent is not actively destroying the illusion, people tend to see complex thinking mechanisms where there are not.

One way to destroy the illusion is predictability, which is a classic flaw for characters in video games. Characters doing over and over the exact same thing are rapidly categorized as being artificial. Adding some unpredictability can better a lot believability (Bryant and Miikkulainen 2006). The difficulty is that too much unpredictability can give the feeling of randomness which can harm believability too. Another flaw is obvious failure to learn. An agent keeping the same behavior even if it is clearly counter-productive breaks the illusion of being "intelligent". That is the main reason why we think that believable agents must be able to learn (Gorman and Humphrys 2007).

The last criterion we will cite makes the difference between believability and realism. Unlike realistic agents, believable agents might be forced to overdo for observers to understand what they are doing (Pinchbeck 2008). Although it can seem strange, it is sometimes mandatory to exaggerate some of the agent's characteristics so that people believe in it. This technique is very often used in arts, especially in cartoons. This means that human-like agents could have a quite low believability. There are however links between realism and believability so it should be a good start to draw inspiration from realism.

Knowing what influence believability, we can design an evaluation method for believable agents. As the cited criteria are quite general, more domain-specific ones should be taken into account in the evaluation.

Evaluation of Believability

As believability is subjective, evaluation is a critical and complex step. Even if it was not intended to, Turing's test (Turing 1950) is still considered as a reference for believability evaluation. In its standard interpretation, a judge must chat with a human and a machine using only text. If, after a certain amount of time, the judge cannot tell which one is artificial, the machine is said to be intelligent. This test's goal was to assess intelligence but it has been much criticized (Searle 1980; Hayes and Ford 1995). This critique, however, does not apply to believability and it is even a very good basis for assessing believability as we defined earlier.

There are many parameters for believability evaluation methods (Mac Namee 2004; Gorman et al. 2006; Livingstone 2006). The first one is to cast or not to cast doubt on the nature of the presented character(s). This choice is often linked with mixing agents and humans so that the judges assess real humans' believability too. This can be useful to avoid bias induced by prejudices and to have a reference: humans usually do not score a perfect believability. Another parameter is the number of questions and answers. Turing's test features only one question and a yes/no answer whereas other tests feature many questions and scales to answer. The former choice may be too restrictive while the latter may result in too much undecided answer to "beat" the test. Another problem is the computation of the overall believability score which, in case of multiple questions, may give experimenters too much influence on the results. To add more objectivity, it is possible to have relative scoring instead of absolute: the score given to a example can answer to "is example A better than example B?". It is also necessary to decide if judges are players or only spectators. While players can actively test evaluated characters, spectators are more focused on them and can notice much more details in the behaviors. Finally the choice of the judges is really important. Cultural origins (Mac Namee 2004) and level of experience (Bossard et al. 2009) may have a noticeable impact on believability scores.

Links between Believability and Video Games

Following our definition of believability, multiple players should be able to play together in the same video game. Indeed, players must not know *a priori* if what they perceive is due to another player or to a program. Believable agents should be used to replace a missing player, having the same role as the others so that we can achieve high level of believability. Of course, single player games can achieve human-like agents but they will not be believable as we defined it.

The video games' complexity of interactions has a really important impact on the believability. If interactions between the players, agents and the game are few, it may be hard to assess believability. For example in a pong game, players can only move paddles so it may be difficult for them to guess the thoughts of the adversary. On the contrary, if players and agents have a wide range of possible actions, they can better perceive the others' thoughts making believability harder to achieve. Video games featuring a complex world and allowing a lot of interactions should be preferred to develop and evaluate believable agents.

Video games are often categorized in different groups. For a testbed, we need a multiplayer game with a lot of possible interactions between players. This kind of games can be found in the action, role playing, adventure and sport games categories. From a technical point of view, adventure and sport games tend to be difficult to modify and in particular to add agents. Role playing involves a large part of communication and natural language is far too much complex to make conversational agents truly believable. Action games are often good choice because they have a quite limited set of actions, simplifying the learning, but their combination are complex enough to be a challenging problem for believability.

The classical example of such action games is first person shooter games. In those games, each player or agent controls a unique virtual body and sees through its eyes. The character can, non-thoroughly , grab items (weapons, . . .), move (walk, jump, . . .) and shoot with a weapon. Each character has an amount of hit points, also known as life points: each time an actor is hit by an enemy fire, a certain amount of hit points are subtracted to the current value. When hit points reach zero, the character "dies" but can usually reappear at another place in the game's environment. Although the concept can seem very basic, it can prove challenging to design believable agents. To make things simpler at the beginning, it is possible to avoid cooperative rules as tactics used may be quite hard to learn.

AGENTS' BEHAVIOUR MODELLING

The general principle of an agent is to perceive its surroundings, take some decisions according to those perceptions and its internal state and then make some actions which depend on the decisions. This is named the perception, decision, action loop because the sequence is repeated over and over. The sets of actions and perceptions are defined by the video game so designing an agent is "only" to find a good mapping between perceptions and actions.

The general structure of the agents' thinking mechanisms is called a model. The model, when parametrized, generates behaviors: it gives a sequence of actions when fed with a sequence of perceptions. In other words, the model is the way to express behavior and the parameters are the behavior.

There are many different agents' behavior models and some can achieve better believability than others. Three domains are very influential on their design: computer sciences, psychology and neurosciences. Computer sciences see agents' models as programs, with inputs and outputs and composed of mathematical functions, loops, etc. Models influenced by psychology try to apply what science understood about humans' behaviors. Finally, models based on neurosciences try to code an artificial brain composed of neurons grouped in different specialized zones. There are many hybrid models and even some of them combines all the three.

There are also two different approaches in the way the agents think: reactive or deliberative. Reactive agents map directly actions to perceptions. The problem is that they tend not to act with long term objectives. Deliberative agents, on the other hand, try to predict the results of their actions. They make plans so that the sequence of their actions will satisfy their needs. Their main weakness is that they are often not very good in rapidly changing environments. Some agents combine reactive and deliberative abilities to have both advantages without the weaknesses.

Models in Industry

In the video game industry, models are often quite basic but can achieve good believability in the hands of skilled game designers. Finite state machines (FSM), hierarchical FSM and behaviors trees are widely used. Those models can express logic, basic planning and are good at reactive behaviours. Another advantage is that those models are deterministic, in other words we can predict what can happen and then try to avoid problems. Their major weakness is that their expressiveness is limited and explaining complex behaviors can become unmanageable.

Another type of model which is used in video game industry is planners. They can give very good results in terms of believability, an example is the game F.E.A.R. which is partially based on the STRIPS planner (Fikes and Nilsson 1971). Agents featuring planners are able to set up complex tactics which have an important impact on believability. However they are not yet capable of learning so they can become predictable. Other models are, by less, used in video games.

Models in Research

In scientific research, many different models are used for agents. Some of the main architectures are Soar (Laird et al. 1987), ACT-R (Anderson 1993) and BDI (Rao and Georgeff 1995). However, these models are not design for believability but for reproducing some humans' thinking mechanisms. Some models are, from the beginning, designed for believability like those used in (Lester and Stone 1997; Mac Namee 2004) and in the Oz project (Bates 1992; Reilly 1996; Loyall 1997) but they are not suited for action games but for emotions and social interactions.

For action games, some work has been done and different models have been presented. (Le Hy et al. 2004) is a Bayesian model which seems to produce quite believable behaviors but lacks of long-term goals. (Robert and Guillot 2005) is a model using classifiers with a online unsupervised learning algorithm but the agents' believability has not been evaluated. (Gorman et al. 2006) use Bayesian imitation, partly based on (Rao et al. 2004), and have a good believability. However it only produces a part of the whole behavior needed for the game. (Gorman and Humphrys 2007) is "only" based on neural networks but due to its learning algorithm, it gives really good results in terms of believability. Here too, it produces only a part of the behavior.

A good choice seems to have an hybrid reactive-deliberative agent as planning has a great impact on believability so as reactivity (Laird and Duchi 2000). Contrary to what is commonly used in video games, a non-deterministic approach is preferable because it can increase believability by avoiding repetitiveness. Agents must be able to learn and, in the best case scenario, should need minimal *a priori* knowledge. This implies that the model does not need to be easily understandable and adjustable as most of its knowledge could come from learning.

Actions and Perceptions

The choice of the model's inputs and outputs is very important because it has a great influence on the final result. What seems logical is to have the exact same sets for the agents and for the players so that agents have all the necessary tools to be believable. However, the perceptions and actions information follow a complicated process between the players' brain and the game. Moreover, having human-like interaction can make the problem even more complicated and instead of improving believability, it could make it worse.

The perception information begins in the video game, each object in it having a set of characteristics like position, rotation, color or texture. All those objects are rendered so they are a sequence of pixel. Those pixels are displayed by a peripheral like a screen or a video projector. Then the information go to our eye and then to our brain. We use visual information as an example, it is also true with sounds, touch and will be true for odors and tastes in the future.

As we can see, the perception information takes multiple forms and all may not be usable or useful for believable agents. From the moment the game's environment is rendered, we enter the domain of artificial vision, audition, etc. Unfortunately, this is a very difficult problem and very few information can be obtained this way. As a consequence, there is a great loss of available information when the rendering is done. The best compromise is to use information directly from the game. It results in having both very precise information, like position, and unavailable information, like lighting.

For the actions, they begin in our brain, then muscles, go through the game's peripherals and arrive in the game. It is possible to simulate peripherals' activity, however it has some drawbacks too. Peripheral information are very low-level so we must translate high-level goals to low-level sequences. If the agent wants to go to some place, it has to find a sequence of peripheral-like actions instead of defining a trajectory for its virtual body.

LEARNING BY IMITATION

Why Imitation?

Form the beginning, we have discussed only about learning without specifying the "how". The reason we choose to focus on imitation only is because of believability. As we defined it, believability means "to look like a player". One efficient way to achieve this is to copy or imitate him/her (Schaal 1999).

Definition and Approaches

Imitation has quite a pejorative connotation as the "innovate, don't imitate" sentence shows. It is however a clever way of learning and very few animals can use it (Blackmore 1999). Moreover, humans use it very often, such often that they do not always notice it and they do it from the very beginning of their lives (Meltzoff and Moore 1977). Imitation is the act of observing a certain behavior and roughly repeating it. This repetition can be deferred and the demonstrator may or may not be aware of being imitated.

From a behavioral and cognitive sciences approach, imitation is the increased tendency to execute a previously demonstrated behavior. The imitated behavior must be new to the imitator and the task goal should be the same as the demonstrator (Schaal 1999). There are some debates on whether or not the strategy should be the same. According to Meltzoff, inferring intentions is necessary for higher level imitation where imitators learn the goals instead of the strategy to achieve them (Rao et al. 2004). This can lead to unsuccessful attempts being use as examples in imitation learning as long as the goal can be inferred. It this case, the strategy may be different from the example given by the demonstrator.

From a neuroscience approach, imitation can be explained studying the brains of evolved animals. Some areas are hypothesized to be involved in imitation learning according to studies on macaques' brains. The quite recent discovery of *mirror neurons* may help in understanding the mechanisms of imitation. Mirror neurons are neurons that fire both when we do something and when we look at someone doing the same thing. However it is not sure that they actually have a major role in the imitation process.

In the domains of robotics and artificial intelligence, imitation learning becomes more and more popular. It has been particularly used in movements imitation in robotics because it reduce the programming time of humanoids robots (Gaussier et al. 1998). Indeed, when we explain a movement to somebody else, showing the example is far more productive than explaining only with words. The same goes with robots and agents: imitation reduce the space of hypothesis the program has to search.

Imitation Learning in Computer Sciences

The main reason we chose imitation learning is that believability is quite linked with resembling to the demonstrator, in our case humans. According to some authors, imitation can lead to believable or at least humanoid agents and robots (Schaal 1999; Thurau et al. 2005). This learning technique seems to be both fast and adequate for our goal.

An interesting model and learning algorithm based on imitation has been designed by Rao and Meltzoff (Rao et al. 2004). What makes it particularly interesting is that it is based on Meltzoff's work on imitation in infants. The model uses probabilities to determine which actions should be taken according to the perceptions. Those probabilities are learnt observing a player's virtual body. Another interesting characteristic is that the agent can guess the player's goals to better understand his/her behavior. This model and algorithm have been extended and used in a action game (Gorman et al. 2006) leading to interesting results.

Some work focused more on the imitation of movement in games which is often highly reactive and strategic at the same time. In their work, Thurau *et al.* used an algorithm named neural gas to represent the geometry of the game's environment (Thurau et al. 2004). They then use an algorithm to learn potential for each neuron/waypoint by giving path followed by players. The sum of those potentials form a field force, attracting the agent to interesting places.

Another interesting work does not use a new model or a new algorithm (Gorman and Humphrys 2007) and it only covers one behavioral aspect: aiming and shooting in an action game. The model consists of three neural networks, one for choosing a weapon, one for aiming and one to choose to fire or not. Those neural networks are trained with a Levenberg-Marquardt algorithm on players' data previously gathered and treated. The results are very encouraging as the agents were capable of short-term anticipation and acted in a really human-like fashion. Agents even copied behaviors due to the use of a mouse: right-handed players have more difficulties to follow targets travelling from the right to the left with their cursor.

One last work based on imitation learning which seems to give very good results is (Le Hy et al. 2004). For basic motion control it use the rule of succession to find probabilities of action by counting the number of occurrences. This rule's main interest is that it estimates the probabilities of actions that did not occur. For the linking of the tasks to achieve, as the model is close to a hidden markov model (HMM), an incremental Baum-Welch algorithm is used to find the parameters. This algorithm permits online learning so it may prove to be a great tool for believability. The main drawback is that we must find a way to identify the current task so that corresponding parameters can be modified.

Learning Constrains

As we said earlier, obvious failure to learn can break the illusion of life. Our first constraint is that the learning process can be observable. This implies that the agent can learn while playing and that it learns fast. The first implication is named *online learning* in terms of artificial intelligence. It means that the model is modified for each given example. It contrasts with *offline learning* where the model is only modified when all examples are given.

Despite the fact that imitation learning can be a fast way to learn, this goal is not easy to achieve. The difficulty in imitation learning is generalization: we do not want the agent to be a film of what the demonstrator showed but to adapt to the environment as the demonstrator did. Learning systems often need examples and counter-examples to learn to avoid over-generalization and over-learning. In imitation learning we only have examples so we might need many of these for the agent to understand how to generalize. This can slow down a lot the learning.

As we saw earlier, learning algorithms and agents' models are heavily linked together. Indeed, the process of learning consists in modifying the model's parameters. The algorithm can even modify the structure of the model. The choice of the learning algorithm or its design cannot be done without choosing or knowing the behavioral model. This makes this choice even more complicated as we must consider the pros and the cons for the couple and not for each one separately.

CONCLUSION

In this paper we outlined the problem of creating believable agents in video games. This problem has three parameters: the players, the video game and the agents. We can add on top the learning mechanisms to make agents capable of evolution. All those factors have an important impact on the final result so their influence should be carefully studied.

First, believability is a very complex notion because of its subjectivity. We defined it as the "illusion that the agent is controlled by a player". Although this definition is quite different from the one used in arts, it makes sense in video games. Indeed, several people can interact in the same game and replacing some of them by computer programs can be very interesting. What makes believability complex is that although players have difficulties in understanding the others' behaviors and often interpret them wrongly, some small flaws can destroy the whole illusion.

The next important factor is the video game itself. A believable agent in an game where it is only possible to chat will be very different from a believable agent in an game where it is only possible to move. Complex games which require both reactive and planning abilities as well as featuring a wide range of possible interactions between players are the most challenging testbeds for researchers.

The agent itself and more precisely the model defining its behaviors is the most obvious factor in the problem. Models can express more or less believable behaviors depending on how they model human characteristics. So, although there are many models both in the industry and the research, they are not all suitable for believable agents.

An agent will not be able to make the illusion last very long if it is not capable of evolving. One possible way of evolution is learning. Imitation learning seems to fit perfectly with the goal of believability. Indeed, it consists in copying the demonstrator acts which is quite close to the notion of believability. It is also a very fast method for transmitting knowledge from one person to another.

The first step in the design of a believable agent is to define an evaluation method. This will help us in knowing which characteristics are important for our agent. This cannot be done without choosing a video game first because the notion of believability depends on the agent's set of actions. It can be also very interesting to evaluate humans' and simple agents' believability first to better understand the problem.

The next step is to choose or design a behavioral model and an imitation learning algorithm. This should be done according to the believability factors discovered in the previous step. Considering the constrains of fast and online learning which make it noticeable, this reduces greatly the available models and algorithms usually used in artificial intelligence.

ACKNOWLEDGMENT

This study has been carried out with the financial support from Région Bretagne.

REFERENCES

Anderson J 1993 *Rules of the mind* Lawrence Erlbaum Associates.

Bates J 1992 The nature of characters in interactive worlds and the Oz project Technical Report CMU-CS-92-200 School of Computer Science, Carnegie Mellon University.

Blackmore S 1999 *The meme machine* Oxford University Press, USA.

Bossard C, Benard R, De Loor P, Kermarrec G and Tisseau J 2009 An exploratory evaluation of virtual football player's believability *in* 'In Richir, S. & Shirai, A. (Eds). Proceedings of 11th Virtual Reality International Conference (VRIC'09)'.

Bryant B and Miikkulainen R 2006 Evolving stochastic controller networks for intelligent game agents *in* 'Proceedings of the 2006 Congress on Evolutionary Computation (CEC 2006)' pp. 3752–3759.

Fikes R and Nilsson N 1971 Strips: a new approach to the application of theorem proving to problem solving *Artificial intelligence* **2**(3/4), 189–208.

Gaussier P, Moga S, Quoy M and Banquet J 1998 From perception-action loops to imitation processes: a bottom-up approach of learning by imitation *Applied Artificial Intelligence* **12**(7), 701–727.

Gorman B and Humphrys M 2007 Imitative learning of combat behaviours in first-person computer games *in* 'Proceedings of CGAMES 2007, the 11th International Conference on Computer Games: AI, Animation, Mobile, Educational & Serious Games'.

Gorman B, Thurau C, Bauckhage C and Humphrys M 2006 Believability testing and bayesian imitation in interactive computer games *in* 'From Animals to Animats 9' Vol. 4095 Springer pp. 655–666.

Hayes P and Ford K 1995 Turing test considered harmful *in* 'International Joint Conference on Artificial Intelligence' Vol. 14 Lawrence Erlbaum Associates LTD pp. 972–977.

Held R and Durlach N 1991 Telepresence, time delay and adaptation *Pictorial communication in virtual and real environments* pp. 232–246.

Laird J and Duchi J 2000 Creating human-like synthetic characters with multiple skill levels: a case study using the Soar Quakebot *in* 'Simulating Human Agents, Papers from the 2000 AAAI Fall Symposium' pp. 75–79.

Laird J, Newell A and Rosenbloom P 1987 Soar: an architecture for general intelligence *Artificial intelligence* **33**(1), 1–64.

Le Hy R, Arrigoni A, Bessière P and Lebeltel O 2004 Teaching bayesian behaviours to video game characters *Robotics and Autonomous Systems* **47**(2-3), 177–185.

Lester J C and Stone B A 1997 Increasing believability in animated pedagogical agents *in* 'AGENTS '97: Proceedings of the first international conference on Autonomous agents' ACM New York, NY, USA pp. 16–21.

Livingstone D 2006 Turing's test and believable AI in games *Computers in Entertainment* **4**(1), 6.

Loyall A B 1997 Believable agents: building interactive personalities PhD thesis Carnegie Mellon University.

Mac Namee B 2004 Proactive persistent agents: using situational intelligence to create support characters in character-centric computer games PhD thesis University of Dublin.

Meltzoff A and Moore M 1977 Imitation of facial and manual gestures by human neonates *Science* **198**(4312), 75–78.

Pinchbeck D 2008 Trigens can't swim: intelligence and intentionality in first person game worlds *in* 'Proceedings of The Philosophy of Computer Games 2008' University of Potsdam.

Rao A and Georgeff M 1995 BDI agents: From theory to practice *in* 'Proceedings of the first international conference on multi-agent systems (ICMAS-95)' San Francisco pp. 312–319.

Rao R, Shon A and Meltzoff A 2004 A bayesian model of imitation in infants and robots *in* C. L Nehaniv and K Dautenhahn, eds, 'Imitation and Social Learning in Robots, Humans, and Animals' Cambridge University Press pp. 217–248.

Reilly W 1996 Believable social and emotional agents PhD thesis University of Birmingham.

Riedl M and Young R 2005 An objective character believability evaluation procedure for multi-agent story generation systems *in* 'Intelligent Virtual Agents' Vol. 3661 Springer pp. 278–291.

Robert G and Guillot A 2005 A motivational architecture of action selection for non-player characters in dynamic environments *International Journal of Intelligent Games & Simulation* **4**(1), 1–12.

Ruttkay Z, Dormann C and Noot H 2002 Evaluating ecas - what and how *in* 'Proc. of the AAMAS02 Workshop on Embodied Conversational Agents'.

Schaal S 1999 Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences* **3**(6), 233–242.

Searle J 1980 Minds, brains, and programs *The behavioral and brain sciences* **3**, 353–382.

Slater M, Usoh M and Steed A 1995 Taking steps: the influence of a walking technique on presence in virtual reality *ACM Transactions on Computer-Human Interaction (TOCHI)* **2**(3), 201–219.

Steuer J 1992 Defining virtual reality: dimensions determining telepresence *Journal of Communication* **42**(4), 73–93.

Thomas F and Johnston O 1981 *Disney animation: the illusion of life* Abbeville Press.

Thurau C, Bauckhage C and Sagerer G 2004 Learning human-like movement behavior for computer games *in* 'Proceedings of the 8th International Conference on the Simulation of Adaptive Behavior (SAB'04)'.

Thurau C, Paczian T and Bauckhage C 2005 Is bayesian imitation learning the route to believable gamebots? *in* 'GAMEON-NA'2005' pp. 3–9.

Turing A 1950 Computing machinery and intelligence *Mind* **59**(236), 433–460.

Weizenbaum J 1966 ELIZA – a computer program for the study of natural language communication between man and machine *Communications of the ACM* **9**(1), 36–45.

DETERMINATION OF INITIAL HIDDEN MARKOV MODELS WITH FEATURE MAPPING: AN APPLICATION TO MMOG PLAYER CLASSIFICATION

Ruck Thawonmas
Intelligent Computer Entertainment Laboratory
Department of Human and Computer Intelligence
Ritsumeikan University
Kusatsu, Shiga 525-8577, Japan
E-mail:ruck@ci.ritsumei.ac.jp

KEYWORDS

Hidden Markov Model, N-Gram, Term Frequency, CRM, MMOG

ABSTRACT

In this paper, we propose a method for determining the model structure and initial parameters of hidden Markov models (HMMs) used for classification of players in massively multiplayer online games. The concept of the proposed method is that of mapping important features of each player type to states of the HMM for that type. Such important features are extracted by N-gram and normalized term frequency, used in natural language processing. According to experimental results, the proposed feature mapping method stably provides high recognition performance and fast learning speed, compared with other existing methods.

INTRODUCTION

The market size of massively multiplayer online games (MMOGs) is growing at a high speed. Although Blizzard Entertainment's World of Warcraft is presently dominating the market, there is still significant competition among other MMOGs, especially those with free subscription but with item-purchasable feature. To keep players in the game, it is crucial that their demands are grasped so that appropriate contents tailored for them can be provided.

In MMOGs, a player can be typically identified according to his/her playing style as a killer, achiever, explorer, or socialiser (Bartle 1996). To exploit such kind of player categorization, customer relationship management for MMOGs should be implemented. Namely, a player is first categorized into one of the pre-defined types based on appropriately selected features from his/her game logs and is then provided a content accordingly. Examples of such contents include those with more hunting opportunities for the killer, a wider variety of collectable items for the achiever, longer-journey missions for the explorer, and higher frequency of social events for the socialiser. This way of content provision should make the player enjoy the game more and hence play longer. We previously showed in a paper (Matsumoto and Thawonmas 2004) that hidden Markov models (HMMs) (Rabiner 1989) trained with player action sequences are effective in classification of MMOG players and that HMMs have higher recognition performance than a variant of k-nn classifier (Thawonmas et al. 2006) trained with player action frequencies. However, the performance of HMMs is in general dependent on their model structure and initial parameters. In our former paper, we empirically used a priori knowledge on the player models of the MMOG simulator in use. In practice, however, it is hard to obtain such knowledge in advance. In this paper, we propose a method for determining the model structure and initial parameters of HMMs by mapping important features of each player type to states of the HMM of the corresponding type. Such important features are extracted by N-gram and normalized term frequency, techniques in natural language processing (Dale et al. 2000).

The rest of the paper is organized as follows. We first describe a brief definition of HMMs and propose a method for mapping important features to HMMs. Then, we describe the MMOG simulator and two player-model types that we use. Next we present our experimental results and end the paper with conclusions and future work.

HIDDEN MARKOV MODEL

The HMM is a tool for statistically modeling a process that varies in time. From the set of observations, time structures hidden therein are derived. Initial applications of HMMs were mainly in speech recognition (Rabiner 1989). More recent applications include bioinformatics (Durbin et al. 1998), MMOG player categorization (Matsumoto and Thawonmas 2004), and musical score following (Pardo and Birmingham 2005).

An HMM can be specified by

1. the set of states $S = \{s_1, \ldots, s_H\}$,
2. the transition matrix A whose elements a_{ij} represent the probability to go from state s_i to state s_j,
3. the set of observation symbols $V = \{v_1, \ldots, v_O\}$,

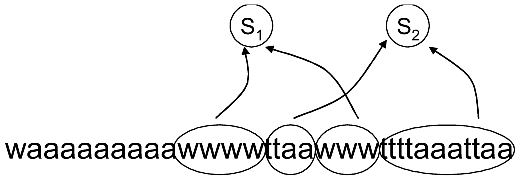


Figure 1: Relationship between N-gram words and model states

4. the emission matrix B whose elements b_{jk} indicate the probability of emission of symbol v_k when the system state is s_j,

5. the set of initial state probability distribution $\Pi = \{\pi_1, \ldots, \pi_H\}$ whose elements π_i represent the probability for s_i to be the initial state.

In our work, the Baum-Welch algorithm (Rabiner 1989) is used to train HMMs, one model for each player type, with a training data set. The training data set consists of both the action sequence and the type of each player. In practice, information from player questionnaires or from game masters could be used for labeling the type of a player in a given training data set; another approach is to perform automatic clustering of players into multiple groups and labeling each group according to its characteristics unfolded by a visual analysis tool (Thawonmas and Hata 2005, Thawonmas et al. 2007, Thawonmas and Iizuka 2008).
To identify the type of an unknown player, the Viterbi algorithm (Rabiner 1989) is used. Inputted by the action sequence of an unknown player, this algorithm computes the log probability of each trained HMM. The unknown player will be labeled by the type of the trained HMM with the highest log probability. As mentioned in the previous section, player-type information can be used in customer relationship management.

FEATURE MAPPING

For determining the model structure and initial parameters of HMMs, we use N-gram and normalized term frequency (NTF), a variant of term frequency (TF). The underlying concept is that of mapping important features of each player type, extracted from the training data set based on N-gram and NTF, to the initial structure of the corresponding HMM. Our basic idea behind this is that, while playing the game, a player stays in a particular state for a certain period and performs related actions, and that states can represented by N-Gram words (Fig. 1).
For player type t, we define an important feature as an N-gram word w with $TF_{w,t}$ beyond a given threshold ρ_t^{TF} and $NTF_{w,t}$ beyond a given threshold ρ_t^{NTF}, where $TF_{w,t}$ is the number of occurrences of w in t divided by the total number of occurrences of all words in t, and $NTF_{w,t}$ is calculated as follows:

$$NTF_{w,t} = \frac{TF_{w,t}}{\sum_t TF_{w,t}} \tag{1}$$

To limit the complexity of resulting HMMs, features that include those with smaller N, extracted previously, are not considered. In addition, a word is defined by symbols and their frequency, regardless of the order of them; for example, "abb" and "bab" are considered the same word. For each player type, a feature is mapped to a state of the corresponding HMM, and the frequency of a symbol in the feature is used directly as the initial probability of emission of the symbol from that state.
The algorithm of the proposed method is given below as follows:

Step 1 - Initialize ρ_t^{TF}, ρ_t^{NTF}, and a word list of each player type t, and set $N = 2$.

Step 2 - Run N-Gram and flag N-gram words with $TF_{w,t} < \rho_t^{TF}$ for each player type t.

Step 3 - For each player type t, select from the unflagged N-gram words those that satisfy $NTF_{w,t} \geq \rho_t^{NTF}$ and do not include any smaller-size words selected in the previous iterations; then add them to the word list of type t.

Step 4 - If no N-gram word has been selected for each player type in the current iteration and at least one of the word lists is not empty, then go to **Step 5**; otherwise, increment N and go to **Step 2**.

Step 5 - For each player type t, add an extra word consisting of all symbols into the word list of type t.

Step 6 - Let H_{max} denote the size (number of words) of the longest word list. For each player type whose word-list size is smaller than H_{max}, repeatedly add an extra word to the word list until its size becomes H_{max}.

Step 7 - For each player type t, construct the corresponding HMM such that it consists of H_{max} states, individually mapped from the corresponding N-gram word in the word list of type t, and initialize each element in the emission matrix B to the frequency of the corresponding symbol in the corresponding word. The initial value of each element in the transmission matrix A and that in the set of initial state probability distribution Π are set to $1/H_{max}$.

MMOG SIMULATOR

Because of difficulty in obtaining real MMOG game logs from game companies, we used an MMOG simulator

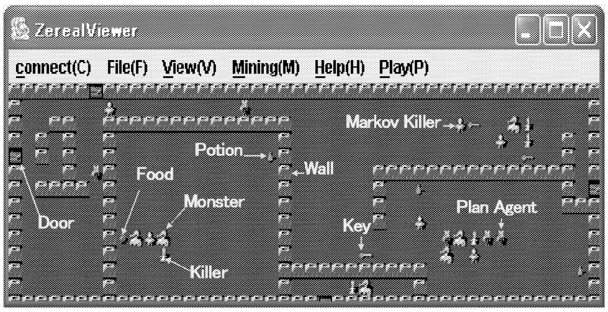


Figure 2: Screenshot of ZerealViewer

called Zereal (Tveit et al. 2003). Zereal is a Python-based multiple agent simulation system running on a PC cluster system and consists of one master node and multiple world nodes. The master node collects the current status (world model) of each world and forwards this information to a client computer for visualization, by a tool called ZerealViewer (Fig. 2), and/or data analysis. A world node simulates all objects such as player agents and monster agents. Other objects include food items and potion items for recovering stamina, and key items for opening a door in order to leave the current world.

In this work, we focused on two types of player agents, Killer and MarkovKiller because they behave like Bartle's killer and achiever, respectively. Each player type has nine actions, i.e., 'walk', 'attack', 'chat', 'pickuppotion', 'pickupfood', 'pickupkey', 'leaveworld', 'enterworld', and 'removed'. The action 'removed' is outputted to game logs when a player agent (or a monster) dies due to its hit point having reached zero. Killer and MarkovKiller have different characteristics, as described below.

- **Killer (K)** mainly pursues the closest player or monster and kill it.
- **MarkovKiller (MK)** selects the next action from multiple candidates using a given Markov matrix.

EXPERIMENTAL RESULTS

In our experiments, we generated game logs by running ten independent Zereal games with 300 simulation-time cycles. Each game consisted of 16 worlds, in each of which there were 50 player agents of each type (**K** and **MK**), 50 monsters, and 50 items for each game object (food, potion, and key). The total agent number of each type per game was thus 800. Next we transformed those raw game logs into action sequences.

We conducted experiments with HMMs whose model structure and initial parameters were determined by

- **Feature Mapping** proposed in this paper, where ρ_t^{TF}, and ρ_t^{NTF} were set to 0.03 and 0.6, respectively,
- **Pruning BIC (Equivalent)** proposed by Bicego, et al. (Bicego et al. 2003) where HMMs initially having ten states were pruned to a smaller size, with the minimum number of states set to 2, using Bayesian Information Criterion, and all elements in the emission matrix B were equivalently initialized with 1/the number of states,
- **Pruning BIC (Random)** which differs to Pruning BIC (Equivalent) in that each element in the emission matrices B was initialized by a random value between 0 and 1,
- **A Priori** used previously in our previous work (Matsumoto and Thawonmas 2004) where the number of states was set to eight and the emission matrix B was initialized with the matrix (Tab. 1) manually derived from the Markov matrix used for implementing MK agents in Zereal. Although this method cannot be used in practice due to unavailability of such a priori, the performance of this method was used as a benchmark index to evaluate the recognition performance of the other methods above.

For all methods, all elements in the transition matrix A and all elements in the set of initial state probability distribution Π were equivalently initialized with 1/the number of states.

The recognition rate of an HMM for each game was derived by averaging the recognition rates when the game logs from worlds 1-8 and worlds 9-16 were respectively used for training and testing the HMM (whether or not it correctly labeled an unknown data to either **K** or **MK**), and vice versa. Our experimental results show that, in terms of the recognition rates of the HMMs over 10 games (Fig. 3), the HMMs initialized by Feature Mapping (examples of them shown in Fig. 4) are comparable to those initialized by Pruning BIC (Random) as well as A Priori, and superior to Pruning BIC (Equivalent). A statistical test was conducted leading to the results that the differences in the performances of Feature Mapping, Pruning BIC (Random), and A Priori were statistically insignificant, and that the performance difference between Feature Mapping and Pruning BIC (Equivalent) was statistically significant.

Focusing therefore on Feature Mapping and Pruning BIC (Random), we found that the former method outperforms the latter one in terms of the learning speed, i.e., the number of iterations for training HMMs until convergence (Tabs. 2 and 3, where each element was derived by averaging the results among ten games).

Table 1: Initial parameters of B in A Priori

State \ Action	w	a	t	p	f	k	e	l	r
fight	0.00	0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.25
talk	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
hunt	0.70	0.00	0.00	0.10	0.10	0.10	0.00	0.00	0.00
transit	0.75	0.00	0.00	0.00	0.00	0.25	0.00	0.00	0.00
go for powerup	0.00	0.00	0.00	0.50	0.50	0.00	0.00	0.00	0.00
flee	0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25
bored	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
transported	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.50	0.00

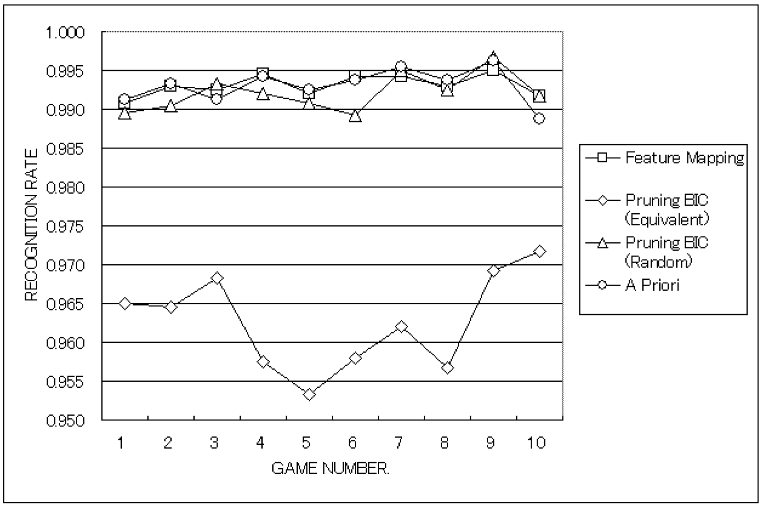


Figure 3: Recognition rates of the HMMs initialized by four methods

CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a method called Feature Mapping for determining the model structure and the initial parameters of HMMs used for classification of players in MMOGs. The HMMs initialized by Feature Mapping have a recognition performance comparable to those initialized by Pruning BIC (Random) as well as A Priori, used as a benchmark index for evaluation of the recognition performance, and superior to those initialized by Pruning BIC (Equivalent). In addition, the HMMs initialized by Feature Mapping could be trained with less number of iterations than those initialized by Pruning BIC (Random). These results indicate that Feature Mapping is more promising than the other existing methods evaluated in the paper.

As our future work, we plan to apply our findings to real MMOG game logs from an edutainment multiplayer online game called The ICE, under development at our research laboratory (Thawonmas et al. 2007, Thawonmas and Iizuka 2008).

ACKNOWLEDGMENTS

This work was supported in part by Grant-in-Aid for Scientific Research (C), No. 20500146, the Japan Society for Promotion of Science. The author wishes to thank Yoshitaka Matsumoto who conducted part of this work for his Master's thesis and thank the anonymous reviewers for their invaluable comments.

REFERENCES

Bartle R., 1996. *Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs.* In *The Journal of Virtual Environments.* 1(1), May.

Bicego M.; Murino V.; and Figueiredo M., 2003. *A Sequential Pruning Strategy for the Selection of the Number of States in Hidden Markov Models.* In *Pattern Recognition Letters.* 24(9–10), 1395–1407.

Dale R.; Moisl H.; and H. Somers E., 2000. *Handbook of Natural Language Processing.* Marcel Dekker.

Durbin R.; Eddy S.; Krogh A.; and Michison G., 1998.

Table 2: Comparison in the number of states and the training iterations among the four methods for Killer

	Number of States		Learning Speed	
	–AVE–	–STD–	–AVE–	–STD–
Feature Mapping	3.45	1.00	6.65	2.35
Pruning BIC (Equivalent)	2.00	0.00	27.00	0.00
Pruning BIC (Random)	4.90	1.94	182.10	84.21
A Priori	9.00	0.00	3.45	0.51

Table 3: Comparison in the number of states and the training iterations among the four methods for MarkovKiller

	Number of States		Learning Speed	
	–AVE–	–STD–	–AVE–	–STD–
Feature Mapping	3.45	1.00	53.55	33.31
Pruning BIC (Equivalent)	2.00	0.00	27.00	0.00
Pruning BIC (Random)	5.65	2.54	180.80	103.95
A Priori	9.00	0.00	3.00	0.00

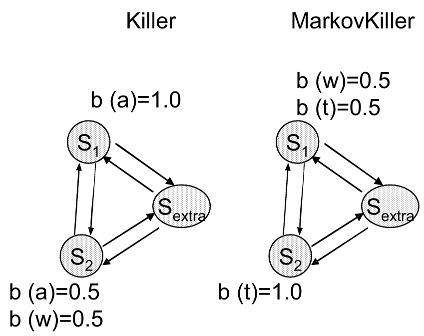


Figure 4: Typical initial model structure and parameters of Killer and MarkovKiller, where $b(x) = 0.11$ for all actions x in the extra state

Biological Sequence Analysis, Cambridge, U.K. Cambridge Univ. Press.

Matsumoto Y. and Thawonmas R., 2004. *MMOG Player Classification Using Hidden Markov Models.* In *Lecture Notes in Computer Science.* Vol. 3166 (Proc. of the Third International Conference on Entertainment Computing, ICEC 2004), Matthias Rauterberg (Ed.), Sep. 2004, Eindhoven, The Netherlands, 429–434.

Pardo B. and Birmingham W., 2005. *Modeling Form for On-line Following of Musical Performances.* In *AAAI 2005.* 1018–1023.

Rabiner L., 1989. *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition.* In *Proceedings of the IEEE.* Vol. 77, No. 2, Feb, 257–286.

Thawonmas R. and Hata K., 2005. *Aggregation of Action Symbol Subsequences for Discovery of Online-Game Player Characteristics Using KeyGraph.* In *Lecture Notes in Computer Science.* Fumio Kishino et al. (Eds.), Vol. 3711 (Proc. of IFIP 4th International Conference on Entertainment Computing ICEC 2005), Sep. 2005, Sanda, Japan, 126–135.

Thawonmas R.; Ho J.; and Matsumoto Y., 2006. *User Type Identification in Virtual Worlds,.* In *Agent-Based Modeling Meets Gaming Simulation.* Springer, (Post–Proceedings of the Session Conference of the ISAGA, International Simulation and Gaming Association, 2003), Series: Springer Series on Agent Based Social Systems, Vol. 2 Arai, Kiyoshi; Deguchi, Hiroshi; Matsui, Hiroyuki (Eds.), Mar, 79–88.

Thawonmas R. and Iizuka K., 2008. *Visualization of Online-Game Players based on Their Action Behaviors.* In *International Journal of Computer Games Technology.* Article ID 906931, 9 pages, doi:10.1155/2008/906931.

Thawonmas R.; Kurashige M.; and Chen K.T., 2007. *Detection of Landmarks for Clustering of Online-Game Players.* In *International Journal of Virtual Reality.* Vol. 6, No. 3, 11–16.

Tveit A.; Rein O.; Iversen J.; and Matskin M., 2003. *Scalable Agent-Based Simulation of Players in Massively Multiplayer Online Games.* In *Proc. of the 8th Scandinavian Conference on Artificial Intelligence.* IOS Press, Bergen, Norway, (SCAI'03), Frontiers in Artificial Intelligence and Applications, Nov.

GAME DESIGN

CREATIVE TECHNOLOGY – THE CTSG: GAME DESIGN IN 7 STEPS

Anton Eliëns [1,2]

[1] Creative Technology, University of Twente
[2] multimedia @ VU University Amsterdam

KEYWORDS
game design, education, workshops

ABSTRACT

This paper reports about a series of workshops in game design, originally meant as promotion for the Creative Technology curriculum of the University of Twente, but evolving into an element in the curriculum itself, with as the goal to create the CTSG, the Creative Technology Superpower Game(s). In the paper, both the goals and motivations of the workshop is dealt with, and the structure and means of executing such workshops, to give potential instructors of such workshops not only an example but some guidance and words of advice as well.

INTRODUCTION

To attract students to our Creative Technology curriculum[1], a series of workshops was organized, with as the main subject *game design*. The workshops were meant to introduce the topics of our curriculum, covering both new media and smart technology, as well as (societal) themes where creative engineering might bring potential solutions. To arouse personal involvement and interest, the participants were asked to state their favorite *superpower*, inspired by a presentation of Matt Costello entitled *Celebrating Creative Genius – Across Media*[2] at PICNIC 2008[3]. These superpowers were later to be used in the actual game design.
Given the fact that our graduates will act, so to speak, as *mediators* in the creative industry and related areas of society[4], as well as the nowadays self-evident requirement to be able to sell your work in pitches and presentations, the overall motto for the workshops was emphasized to be:

communication is the key to creativity[5]

In the course of the workshop(s), students were required to give presentations ranging from 1-minute pitches to 5-10 minute competitive proposal discussions. It must be noted that *communication* was also one of the potential game topics, as alienation and loneliness may considered to be potential dangers of our highly technological society, and as such one of the paradoxes of the (socially) networked society. And, as an aside, communication is, as many of us know by experience, generally not the strongest point in CS departments, reason the more to bring the *communication* topic explicitly to the attention of our (future) students.
The material in this paper stems primarily from a re-enactment of the workshop in the *multimedia authoring*[6] class of the author, with more senior students, which was held to have a somewhat more controlled setting in which to obtain materials and ideas. Anyway, in my experience, the challenge of designing a game around superpower(s), or a combination of superpowers, to *improve the world* has proved to be appealing to all levels of students.

structure The structure of this paper is as follows. First an outline of the creative technology promotional workshops will be given, with a brief explanation of why we chose for this format, that is *paper&pencil* based work. The possible themes of game design will be discussed, as well as the structure of the workshops and the actual assignments. Then we will look at the results of the (re-enactment of the) workshop, that is both superpowers and related game scenarios. After discussing, on a more general level, techniques for design and creativity, we will explore the future potential of game development around the superpower theme(s), in particular for the Creative Technology curriculum. And finally, we will draw the conclusions.

CREATIVE TECHNOLOGY PROMOTIONAL WORKSHOPS

A *paper&pencil* workshop was decided for, to be able to cover a lot of ground, so to speak, in a relatively short time, usually between 45 minutes to an hour and a half. The alternative of using computer-based tools to introduce (potential) students to creative work seemed

[1] create.ewi.utwente.nl
[2] vimeo.com/3188577
[3] www.picnicnetwork.org
[4] create-media.blogspot.com
[5] www.cs.vu.nl/∼eliens/.CREATE/create-workshop.html
[6] www.cs.vu.nl/∼eliens/mma

at first more appealing, but would take on reflection simply too much time. Although the proposal for *paper&pencil* based work originally met with some scepticism, the results in terms of interest of the participants and information conveyed to and about the participants proved to be (very) satisfactory. It must be remarked, however, that the workshops were complemented by a hands-on practicum with sensors and arduino boards[7], to give potential students also a feel for the (smart) technologies used in the curriculum. Such a practicum was, however, not part of the re-enactment described in this paper.
The (original) workshops are organized in seven rounds, as indicated below:

workshop game design – in 7 rounds

1. introduction – getting to know eachother
2. selection of game theme(s) – the playground
3. exercise(s) – associative chaining
4. visual style – design & reference(s)
5. narrative content – story line(s)
6. game mechanics – play & score(s)
7. wrap up(s) – speak out!

After a first round, where each participant was asked to state his/her name and favorite superpower, we started with the assignments. However, before starting the actual assignments the participants/students are primed by the presentation of a collection of videos, with brief explanatory remarks, using a selection of clips from .CREATE TV[8], ranging over *future technological innovations*, *urban threats and violence*, *the impact of media*, *novel ways of human-computer interaction*, as well as a selection from *watchmen*[9], to emphasize the personal responsibility of the students to become a *superhero*.
As an observation, in one of the workshops a participant remarked after hearing all participants' superpowers that he felt that he *knows these people better than his own friends*. Another important exercise, by way of warming-up, that is bring the participants in the right mood, is *associative chaining*, in which the assignment is to make a brief story by having each participant contribute a phrase or line, in turn, until the story reaches a certain length, say 20 lines for a group of five. For this assignment, as well as the later assignments, groups of 4-5 people were formed. Dependent on time, a selection was made from the other assignments, focussing on visual style, narrative content or game mechanics. After each assignment, including the *associative chaining*, the participants/groups were asked to give a brief presentation, with a comparative evaluation with a slight competitive edge at the end of the workshop.

[7] www.arduino.cc
[8] www.cs.vu.nl/∼eliens/ximpel/project/tv
[9] watchmenmovie.warnerbros.com

THEMES AND VARIATIONS

An auxiliary goal of the workshops, that is of both the promotional workshops and similar workshops that are part of the curriculum, is to inspire and prepare (potential) students to become creative engineers that are able to apply the new technologies in a societal context, and as we hope, in a way that will *improve the world*, with as an additional observation that ultimately the future world is their world, after all. Six potential topics were indicated and in addition (potential) students are invited to come up with their own ideas, to which the response is usually (very) low!

theme(s)

1. **urban** – (un)safety in urban environment(s)
2. **climate** – reduction of energy consumption
3. **fitness** – (social network) support for sport and fitness
4. **media** – prevention of information overload
5. **scenario(s)** – event(s) in public area(s)
6. **communication** – in private/public space(s)
7. **idea(s)**? – open call for additional theme(s)

Interestingly, in the experimental group at VU, students were extremely reluctant to commit themselves to one of these goals (2 out of 15), whereas in the promotional workshops the (not yet) students eagerly selected one or more of the themes, with an overall preference for (3) and (6).
The format in which these topics were discussed differed widely among the various workshops. Sometimes a plenary brainstorm was held, and sometinmes these topics were discussed in small groups, where the instructors asked the participants to indicate the personal relevance of these topics and urged for potential solutions.

STRUCTURE AND ASSIGNMENTS

After the first round of introduction, the participants of the workshop are divided into groups of 4-5 people, and set to work on designing the game map and/or front portal, define the game mechanics of a mini-game and describe the story-line as well as the actual tasks of the player(s) and associated scores.

structure

1. create (y)our identity – draw (y)our avatar
2. invite the player(s) – design game map / front portal
3. invent game mechanic(s) – sketch scenario(s)
4. create challenge(s) – define task(s) & score(s)

The first assignments, in all workshops, including the re-enactment, is to draw a self-representation, that is an avatar, possibly with attributes to indicate the endowment with superpower(s).

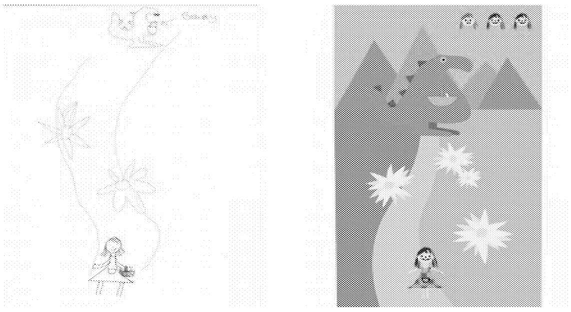


Fig 1. (a) student sketch (b) enhanced version

Some participants, included already a game environment, as for example depicted in figure 1 (a), a sketch originally made by a (potential) student, and 1 (b), which shows an enhanced version of that sketch and explicitly includes the dragon which is hardly visible on the original, made by Marek van de Watering who acted as a co-instructor in one of the early game design workshop(s)[10].
For the (potential) Creative Technology students, it is important to take into account the technical prerequisites, that is the technology that may be used to realize the game.

(pre)requisite(s)

- smart technology – rfid tags, motion sensors, ...
- new media – camera(s), screen(s), game world(s)

Interestingly, these technologies may effectively take the game out of the console or computer terminal into (interactive) space(s), encompassing both the private and public domain. Ultimately, such technologies allow for the realization of ARGs (Alternate Reality Games), that (minimally) enhance existing (urban/corporate) spaces into game environments.

SUPERPOWER & GAME SCENARIOS

The workshop format is a convenient way to involve students in design activities, and bring focus into game design. Although with different format(s), a similar approach was used for the *game @ VU*, *VU @ Second Life*, and *Clima Futura*, see Eliens & Bhikharie (2006), Eliens et al. (2007a) and Eliens et al. (2007b).
In the re-enactment of the workshop at VU University Amsterdam, with more senior (that is pre- and post-graduate level) students, 15 students participated. They were given an explanation of the author's intentions as well as an overview of the structure of the workshop, unlike the participants of the promotional workshops for Creative Technology. The first round resulted in the following list of superpowers:

> teleporting (2), flying/levitation (2), mind reading (2), create what you think, change materials, see ahead in time, shape shifting, invisibility (2), control the force of nature, healing

[10]creativetechnology.eu/wordpress/?p=194

Such a list is representative of most of the workshops, although alternative superpowers, from other workshops, include *devastating smile*, *clearvoyance*, and the *capability to detect invisible*(s).
By way of illustration, how a superpower may be connected to a scenario of playing a minigame, look at the following quote:

> *my superpower ... healing ... to deal with exponential epidemics, my scenario would be a world map with potential sources, and as a minigame I propose one to improve the skill of taking instant decisions of where to go ... and, indeed, practice healing*

The students participating in the re-enactment were explicitly asked to think about how to apply their superpower to improve the world, or more specifically, to think of a mini-game that one way or another was related to their superpower. For this the students were invited, after stating their identity and superpower, to create groups. This resulted in six groups of about 2-3 people.
As already observed above, the students @ VU were extremely reluctant to choose for one of the, admittedly ambitious, game themes, nevertheless they experienced no problem in coming up with a variety of game scenarios, which are (in outline) summarized below.

CTSG – scenario(s)

1. trophee collection game(s) – fly & teleport
2. conflict(s) – cooperation and interaction
3. virtual friend(s) – act as intermediate in learning
4. combat game(s) – invisibility and self transformations
5. social party game(s) – sequence of confrontation(s)
6. labyrinth(s) – material transformations and parallel presence

Below we will briefly characterize the first four scenarios, that are related, respectively, to fig. 2 (a),(b) and fig. 3 (a),(b).

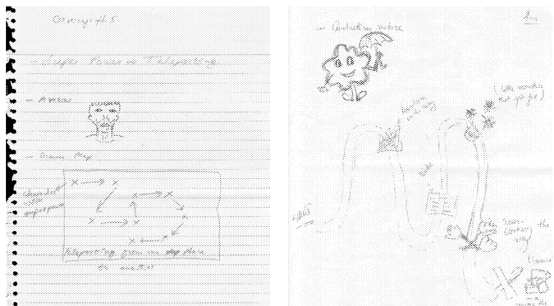


Fig 2. (a) teleport(s) (b) control nature

The first scenario, fig 2. (a), involves *fly* and *teleport* as superpower, and amounts to:

scenario 1: trophee collection

(1) *The character has the ability to fly and teleport.* (2) *He as some amount of energy which limits his abilities of teleport and flying.* (3) *The character should collect items using his superpowers, that will provide him clues to reach his destination.* (4) *The character has to fight with some bad guys, and finally* (5) *fight the main bad guy ...*

The second scenario, fig 2. (b), involves *control of nature*, *create what you think* and also *instant knowledge* (not mentioned in the list) as superpowers. Due to mobility of one of the group members, it bears some relation to scenario 3:

scenario 2: conflict(s)

(1) *The player earns points/credits by helping eachother to solve problems.* (2) *Credits can be used to be goodies or treats. of teleport and flying.* (3) *The player has to take care of* blobs *and keep them happy with goodies and treats.*

The following scenario does not very clearly involve a superpower but as, hardly readable, indicated in fig 3. (a), one of the (desired) superpowers is the ability of *instant learning*, in particular languages. Apart from that, the scenario is interesting, and seems to apply to a wide range of contexts:

scenario 3: virtual friend

I want a virtual friend with whom I can interact and play, but who can act in the same time as my personal assistant, and can collect knowledge for me, and get to know me and what I am interested in, so that s/he can act as an interface between the knowledge and me. You must however earn credits to keep your (virtual) friend happy ...

Finally, as what may be considered a more conventional game scenario, we have a combat situation, fig. 3 (b), which involves the superpowers of *invisibility* and *self-transformation*:

scenario 4: combat

(1) *Superpowers involved are invisiblity and the power to change limbs/weaponry.* (2) *Turn-based fights, with possible avoidance (using stealth), confrontation using matching powers/weapons.* (3) *The mission is to find and detonate a bomb.*

It is interesting to note that some superpowers were added during the development of the game scenarios, and that others became more well-defined during the process of development.

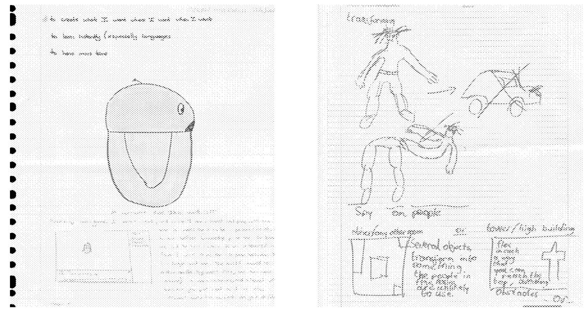


Fig 3. (a) power to create (b) transforming

TECHNIQUES FOR DESIGN & CREATIVITY

Reflecting on issues of design and creativity it is worthwhile to consider, following Jones (1992), whether *design is art, science, or a form of mathematics*. Clearly, our workshops are focussed on the early phases of design and involve, as indicated in the table below, both *experience* and *brainstorming*:

	divergence	transform	convergence
(1)	*experience*	...	value analysis
(2)	brainstorming	...	...
(3)	...	...	dependency graphs
(4)	...	...	*requirements*

(1) community, (2) system(s), (3) products, (4) component(s)

Other methods of design, in particular requirements engineering or more mathematically oriented methods, such as dependency graphs, must considered as detrimental to the ideational phase of design. To assist the instructor we provide a list of techniques to promote/encourage creativity, taken from Jones (1992):

creativity technique(s)

- analogy/metaphor – similarity & figures of speech
- brainstorming – large quantity of ideas in short time
- blue slip(s) – small cards to express ideas
- extrapolation(s) – apply proven methods
- progressive abstraction – alternative problem definition(s)
- 5W+H technique – who/what/where/when/why/how
- force field(s) – identify contributing or hindering force(s)
- peaceful setting – relax and open mental process
- problem reversal(s) – to provide different framework(s)
- association(s) – use inclination to associate things
- wishful thinking – to counteract analytical approach

As a remark, personal experience seems to be a key motivator, and even more so for young (potential) students, an observation that is supported by Eliens (1979) and Munari (1966).

ELABORATION(S) – GAME DEVELOPMENT

In the first year course *we create identity*[11], students were repeatedly confronted with questions such as: *what is the CTSG*?, *why the CTSG*? and *how to realize the CTSG*?, without further explanation of what the phrase *CTSG* stood for. After students discovered the meaning of *CTSG* (Creative Technology Superpower Game), it became clear that, whereas ideas may be easily generated, as testified by the workshops, the realization of such a game may be quite daunting. However, the 2009 group of students succeeded in realizing an impressive exhibit in the newly built SmartXP lab, an interactive space with an abundance of technical facilities for research and education, after which the group gained enough self-confidence to take on the mission of actually developing the **CTSG**, in a range of possible games, covering relatively simple educational games, such as math games, Eliens & Ruttkay (2009), as well as urban games in the context of the GOGBOT festival[12], a regional festival featuring technology and art. As we observed in Eliens & Ruttkay (2008), a very helpful set of criteria for distinguishing games from other (online) applications were presented to us in a workshop on educational games in a museum context, which mentions 4 essential characteristics or criteria to assess the extent whether an application may considered to be a game:

criteria

- *challenge* – relevance, feedback, confidence
- *curiosity* – cognitive or sensitive discrepancy
- *control* – contingency, choice, power
- *context* – intrinsic or extrinsic metaphor(s)

Leaving a more detailed interpretation of *challenge*, *curiosity* and *control* (again) to the inventiveness of the reader, the *context* characteristic, however, needs some elaboration, especially when discussing the **CTSG**s. As we observed, it is rather easy to use extrinsic metaphors or game formats for arbitrary content. For example, a memory game can be reused over and over again, just by changing the images according to the subject, that is language learning, climate change, etcetera. These kind of mini-games or casual games lend themselves to a variety of learning tasks and may be constructed using pre-defined game formats. Much more difficult is to design games with an intrinsic relation to the topic. It seems, however, that the notion of **superpower** provides a suitable metaphor, to relate real-world situations to a game, in which the player may perform actions or solve problems using superpower(s) and thus become a *superhero*, if only by analogy with (superhero) movies. Cf. Bolter and Grusin (2000).

[11] www.cs.vu.nl/~eliens/.CREATE/create-ca1.html
[12] 2009.gogbot.nl

CONCLUSIONS

In this paper we have reported on our experiences with a series of workshops in game design. Despite the promotional nature of these workshops, the material presented here should be of use to instructors who wish to set up similar workshops. As a tool to focus on topics of societal relevance, the format of the workshop allows instructors to get to know their students and motivate them, using the superpower metaphor, to become personally involved in the various aspects of game design.

ACKNOWLEDGEMENT(S) Thanks to Marek van de Watering for his support in setting up the workshop in it's initial stages and to Timen Olthof for administering the experimental setting at VU, and his contribution of actual scenarios. Also thanks to the students at VU who participated in the experimental re-enactment of the workshops, and above all to the students Creative Technology, for (continually) providing the inspiration and motivation, and ultimately the workforce, for the CTSG.

REFERENCES

Bolter J.D and Grusin R. (2000), *Remediation – Understanding New Media*, MIT Press

Eliens A. (1979), Creativity: reflection and involvement, Ninth Int Conf of Aesthetics, Dubrovnic, August 1979

Eliens A. and S.V. Bhikharie (2006), *Game @ VU – developing a masterclass for high-school students using the Half-life 2 SDK*, In: *Proc. GAME'ON-NA'2006*, Monterey, USA

Eliens A. Feldberg F., Konijn E., Compter E. (2007a) , VU @ Second Life – creating a (virtual) community of learners, In *Proc. EUROMEDIA 2007*, Delft, Netherlands

Eliens A., van de Watering M., Huurdeman H., Bhikharie S.V., Lemmers H., Vellinga P. (2007b), Clima Futura @ VU – communicating (unconvenient) science, In Proc. GAME-ON 07, Bologna, Italy

Eliens A. and Ruttkay Z. (2008), Record, Replay & Reflect – a framework for understanding (serious) game play, In Proc. Euromedia 09, Brugge, Belgium

Eliens A. and Ruttkay Z. (2009), Math game(s) – an alternative (approach) to teaching math?, In Proc. GAME-ON 2009, Dusseldorf, Germany

Jones J.C. (1092), *Design Methods*, Wiley (2nd edn)

Munari B. (1966), *Design as Art*, Penguin

INTERACTIVE SPACE(S) – THE CTSG: BRIDGING THE REAL AND VIRTUAL

Anton Eliëns [1,2]

[1] Creative Technology, University of Twente
[2] multimedia @ VU University Amsterdam

KEYWORDS
interactive space, locative game design, education

ABSTRACT

In this paper, ideas will be presented how to realize games or playful activities in interactive space(s), having a real (spatial) component as well as a representation in virtual 2D or 3D space, by means of web pages and/or online games. Apart from general design criteria, the paper discusses a number of sample scenarios resulting from a workshop with students from the Creative Technology curriculum on the theme of *Creative Technology Superpower Games* (CTSG), where the main focus is to investigate playful applications of technology to emulate or support players' superpowers, such as invisibility, telekinesis, etcetera. Although the paper does not intend to provide a methodological framework, we will briefly discuss the various logics underlying spatial and locative games, or more generallly games that may take place in interactive space(s).

INTRODUCTION

Technology increasingly invades our private and public space(s), partly as a result of our own efforts and demands for being informed, entertained and served, that is our supposed convenience, and partly due to causes beyond our immediate personal control, such as security, public transport systems, and other bureaucratic concerns that apparently require an extensive application of monitoring and tracking devices, including cameras, RFID scanners, etcetera.

The complexity of such systems requires what may be called *research in context*, and has led to interdisciplinary research laboratories for research with interactive space(s) as the main theme, across domains such as domestic entertainment, smart workplaces and healthcare. At the University of Twente, we have recently opened the SmartXP facility, with as a mission statement:

smartxp.ewi.utwente.nl

> Research in ubiquitous computing is inherently a multi-disciplinary activity. Creative application and development of technology is one essential ingredient of ubiquitous computing which however requires in-depth access to the relevant technologies that goes beyond a mere literature study or an occasional cooperation.
>
> (...) We are establishing a living environment as an open resource and show-case facility for ubiquitous technology in realistic settings: the Smart eXperience lab (SmartXP).

To be clear, although the environment provides ample technology, research in this area is explicitly meant to be interdisciplinary:

multi-disciplinary research

> The SmartXP is an important tool for prioritising the research agenda, both in the technical and social sciences. The SmartXP provides essential facilities for small and medium scale projects with flexible experimentation grounds in a realistic setting, with knowledgeable users and competing technology.

There is by now a rather long standing tradition in academia to deploy games as a format for doing research in ubiquitous computing as well as other areas demanding a multi- or even trans-disciplinary approach. In our bachelor curriculum Creative Technology, moreover, we have opted for an approach that takes games and playful applications as a core element of the program, as a means to bridge the gap between the primarily technical disciplines of engineering and mathematics, and the exploration and creativity needed for innovation and novel solutions to the technological dilemmas of the twentyfirst century. Relying on the intrinsic motivation(s) and creative aspiration(s) of our students, we find support in the following characterization of creativity:

> creativity = connecting dots one would not look at, normally ... Alain Kay

From a pioneer in innovative computer applications, Alan Kay's original characterization provides a guideline for both the design and realization of systems that deal with interactive space(s), bridging the real and the virtual.

structure The structure of this paper is as follows. We will start with a brief overview of spatial and locative (aspects of) games, after which we will sketch the design space for our exploration in the design of games for interactive space(s). The main body of the paper, then, consists of samples and scenarios resulting from a workhop with our Creative Technology students, followed by a discussion of the underlying logic(s) and game mechanics. We will look at how these scenarios may be used for the CTSG, the Creative Technology Superpower Games, identifying the main challenges, after which we will present some conclusions and indicate directions for future developments.

SPATIAL & LOCATIVE GAMES

In her fascinating book, *Critical Play*, Mary Flanagan relates how artists have explored games and playful applications to bring about awareness of (the) human (living) condition(s), and more in particular, in recent times, how *situationism* critically investigated urban space using mechanisms of *unplay, re-skinning & re-writing*, Flanagan (2009). In addition, an exhaustive inventory of existing games, and their application in critical play, is given, ranging over doll houses, board games (such as Monopoly), language games, as well as more recent performative and locative games, such as C2BK[1].
Another source of inspiration is provided by the interactive new media installations exhibited at cinekid.nl[2] (in Dutch), where cameras, scanners and light sensors were used as interaction devices, allowing children to play with sound, motion graphics and animations.
Apart from design issues and technical possibilities it is, however, before we start our exploration of possible scenarios, worthwhile to consider why people would like to play (...) games, following the **when we play games** blog from the *Ubiquitous Play in the Everyday* class, with world-famous instructor Jane McGonigal, at thismightbeagame.blogspot.com[3]:

values (1)

> When we play games, we experience relaxation, concentration, cohesion, elation, adventurous thinking, constant challenge, and relief. We want more of these things in everyday life.

However, not only positive experiences or motives may induce playing games:

values (2)

> When we play games, we feel awed, sneaky and backwards. We should feel like this in real life, too.

[1]cruelgame.com
[2]www.cinekid.nl/nl/festivalprogram/part/19
[3]thismightbeagame.blogspot.com/2007/02/why-we-make-ubiquitous-game-our.html

Ultimately, games may be considered as empowering the player(s), as a means to improve oneself:

beliefs (1)

> A well-designed game can make someone more responsive to clues in the environment. It can help someone recognize opportunitIes.

PLAYGROUND DESIGN SPACE

The starting point, or impetus, to our explorations was provided by a student in the course *web technology* who, in his turn inspired by google wave(s)[4], created a simple application, using javascript and sqlite, to make collaborative text collages. Extending the data format with a state variable for each item allows for displaying simple narratives or slogan(s)[5], that progress on each movement, as illustrated in fig. 1.

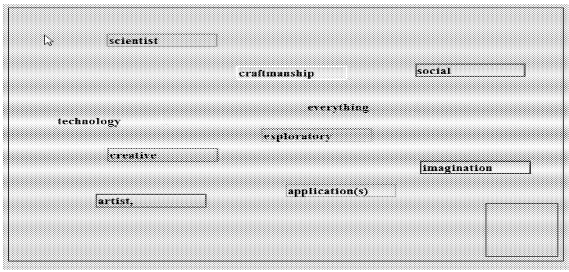


Fig 1. colored story markers

Albeit simple, a two dimensional display, with positional information, may be used for a variety of purposes:

positional marker(s)[6]

- storytelling – undirected token with temporal sequence
- grading – positional markers with interpretation
- voting – place attachments with posts or milestones

For grading, when giving the x and y axes an interpretation, as for example, respectively, level of technical skill and creativity, we may simply map each students' performance on this display and assign a color along these dimensions, with red for creativity and blue for technical skills, using the formula:

positional color(s)

R = (255 * ((height-y)/height)) % 255)
G = ((255-abs(R-B)) * (R/255) * (B/255)) % 255
B = ((255 * (x/width)) % 255)

which effectively assigns black to a student that does not perform at all and white to the overall excellent student.

[4]wave.google.com
[5]www.cs.vu.nl/∼eliens/media/sample-jquery-slogan.htm
[6]www.cs.vu.nl/∼eliens/.CREATE/marker-nm1-09.html

SAMPLE(S) & SCENARIO(S)

Following the format of the workshop(s) game design, as described in Eliens (2010), and taking into account the playground design constraints indicated in the previous section, the students Creative Technology were given the instruction to design a visual representation of the player(s), to develop a playground with a clear relation to (the) real (SmartXP) space, and to write down narratives or scenarios for game play, that involve both the real and virtual space(s). In summary, the assignments were:

assignment(s)

1. avatar – for (self) representation
2. playground – relating real & virtual space(s)
3. scenario(s) – narrative(s) & game mechanics

The participants were strongly encouraged to take the SmartXP lab as the real playground, but were left free what kind of game to develop and their topic of choice, that is a game for pure entertainment, or a serious game, possibly reflecting their own struggle in mastering the technical disciplines of the curriculum. The total time available for the assignments was thirty minutes.
A first look at the results learns that both the playgrounds and scenarios are highly divergent, which may be due to the allowed free choice of the topics. Nevertheless, despite this diversity, a closer look at the proposals produced by the students is instructive, perhaps not for the realization of a final game, but as a first exploration of the design space for a playground encompassing real and virtual space(s).

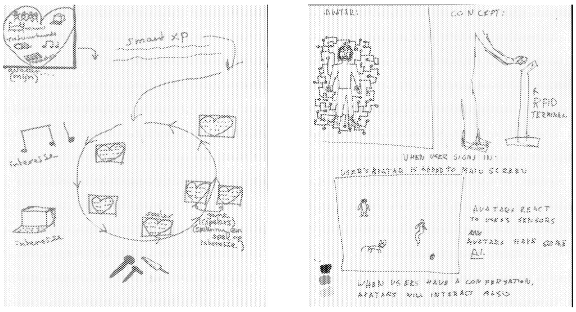


Fig 2. (a) physics & stuff (b) (rfid) identity

The first sample, fig. 2(a), takes physics as a topic, and features as tags

sample 1

physics & stuff – *SmartXP interest avatar physics miauw*

with the final tag *miauw* (representing the sound of a cat), as an obviously playful or nonsensical element.
The second sample, fig. 2(b), involves both avatars and rfid terminals:

sample 2

(rfid) identity – *when user signs in user's avatar is added to main screen, avatar reacts to user's sensors, avatars have some A.I., when users have a conversation avatars will react also.*

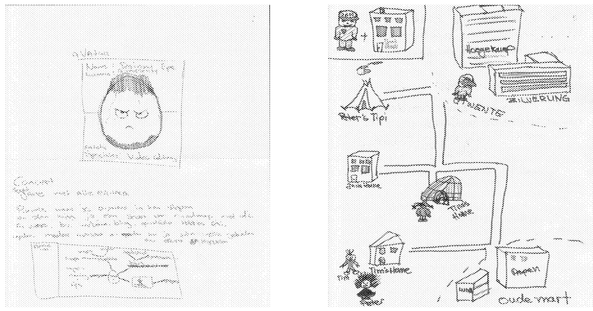


Fig 3. (a) personal avatar(s) (b) dynamic city

The third sample, fig. 3(a) is centered around avatars with personal skills, and proposes a sports game:

sample 3

personal avatar(s) – *e.g. with skills (video editing); concept: sport game with all avatars; space where avatars can be dragged to, resulting in a mind map of that avatar, e.g. with a website, portfolio, blog, hobbies, etc.*

In sample 4, fig. 3(b), a city is taken as the location of activity:

sample 4

dynamic city – *where everybody has a house and an avatar: People's houses get integrated into the city. Everybody can update his position in the city according to his position in real-life. (possibly with RFID scanners instead)*

Note that updates in position are realized using RFIF scanners.

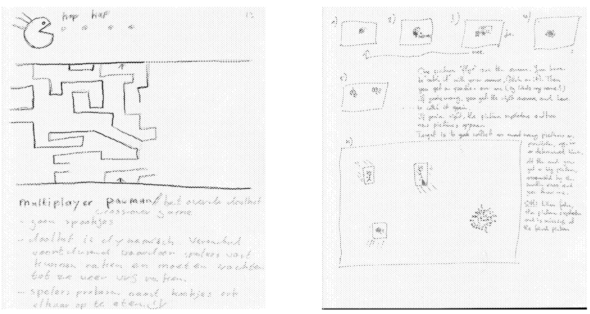


Fig 4. (a) pacman labyrinth (b) catch the image

Both sample 5, fig. 4(a), and 6, fig. 4(b), may be characterized as entertainment or action games, with in sample 5 an interesting, though slightly morbid, twist.

sample 5

pacman labyrinth – *multiplayer pacman / haunted labyrinth / cross-over game; no fairytales! labyrinth is dynamic. It changes continuously, so that players get stuck and have to wait until they are released. Apart from eating cookies, players also eat eachother*

How to translate encounters between players to real space is not entirely clear, unfortunately.
Sample 6 may, moreover, be interpreted as a puzzle game, to uncover the global picture:

sample 6

catch the image – *one picture "flies" over the screen, you have to "catch it" with your mouse (click on it). Then you get a question about me (e.g. what is my name?). When you're wrong, you get the right answer and have to catch it again. When you're right, the picture explodes and two new pictures appear. Target or goal is to catch as many pictures as possible, e.g. in a determined time. At the end, you get a big picture assembled by the smaller ones, and you have me!*

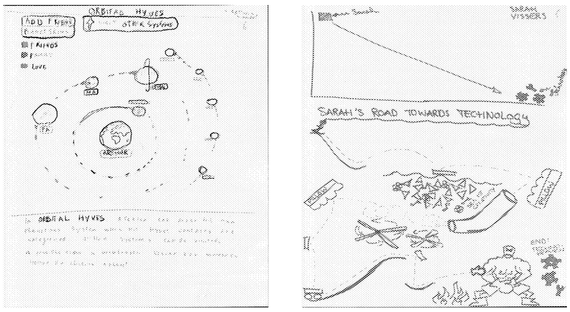


Fig 5. (a) orbital hyves (b) road to technology

Sample 7, fig. 5(a), uses space to visualize a social network, or more in particular, the proximity of relations expressed by social connections:

sample 7

orbital hyves – *friends / family / love; a person can make his own planetary system, where his hyves contacts are categorized, other systems can be visited. You can view someone's house by clicking a planet.*

Sample 8, fig. 5(b), takes again the study trajectory as a topic, which might be mapped to real space as a sequence of hordes, that is obstacles to take, possibly in competition with fellow players/students:

sample 8

road to technology – *the idea is quite simple: a sketch of my interpretation of getting towards the required knowledge to be a great creative technologist. It turned into some sort of a "classic" Mario-ish game in which I am facing different obstacles.*

It would be nice to have our entire class to enter this game, some floating quicker than others and people perhaps bumping into eachother, creating more speed.

As a final scenario, for which no suitable drawing was produced, look at the following story which takes the original suggestion of progress in the study along the dimensions *creativity* and *technology* as its theme:

story of SmartXP

SmartXP – 2 planes [creativity / technology]

SmartXP is divided in two planes, full with a variety of objects. Objects in the Creative part look at first sight creative, but are deep inside still Technologic. For objects in the Techology part, it is the other way around.

A person walks around, observes, and uses. After that s/he goes to the CreaTester, which tests the person on what s/he observed in the objects, what choices s/he made, etcetera.

The system computes the location of the person, and beams it on the screen.

The *story of SmartXP* does not only take the constituent dimensions of Creative Technology as its theme, but does also provide opportunities for critical play, in encouraging the player to discern the creativity underlying technological applications but, on a deeper level, to discover the technological substratum underlying all creativity.

SPATIAL GAME LOGICS

In *Expressive Processing*, Noah Wardrip-Fruin provides a thorough analysis of the computational processes underlying games and interactive storytelling. He observes that most classic games, including Pong as well as adventure games, are based on spatial logics, that is governed by the dynamics of navigation, motion and collision detection, Wardrip-Fruin (2009). He introduces the notion of *textual machines* to counter-act the dominance of spatial logics and to advocate what he calls *lyric engagement*, that is more closely related to the experience of music, as opposed to physical manipulation and visual experience in for example action games.
As indicated in the introduction and the section(s) delineating the design space of our interactive playground(s), our challenge in realizing interactive space(s) is to find suitable mappings between the symbolic space of for example learning trajectories and spatial representations, where we may freely use basic game play mechanics

such as motion and collisions, to represent challenges such as mathematical puzzles, language games or role changes in strategy games. In the realization of the game(s), the player must be able to understand such referential jumps, without being interrupted in game play. Cf. Bolter and Grusin (2000).

THE CTSG – WORKFLOW(S)

In the aforementioned new media installations at cinekid.nl, one stood out because of its underlying metaphor, which may be expressed as:

metaphor(s)

I am a pencil, and I walk the earth ...

characterizing a navigation tracker, using GPS.
In our effort to create games for interactive space(s), we may rely on our previous work when facing subtasks such as *replicating real environments*, cf. Eliens & Bhikharie (2006), Eliens et al. (2007a), *providing charts, maps or blue print of regions*, cf. Eliens et al. (2007b), Eliens et al. (2008), or *defining trajectory of repeatable challenges*, cf. Eliens & Ruttkay (2008), Eliens & Ruttkay (2009).
Our main technological efforts, primarily in terms of development and maintenance, may be summarized as:

technology exploration(s)

- identity – RFID, image analysis
- track individual change(s) – data management

Additional technologies that are worthwhile to look at are, among others those used in geodart[7] that allows for creating games about geographical trivia using embeddable flash maps[8], a technology that for the same token may be used for one of our voting games, or for visualizing events that occur in our interactive space(s)! Our main challenges, however, are to find a coherent scheme in which to fit the various ideas, and to develop a framework that allows for the adaptation of our game(s) to the particular usage context of (y)our interactive space(s).

CONCLUSIONS

In this paper, we have sketched our ideas and sample scenarios for games in interactive space(s), within the ongoing theme of game development in the Creative Technology curriculum. Although a significant part of the development effort remains unexplored, the material presented here should clarify issues involving mappings between real and virtual spaces, as well as metaphors and logics suitable for dealing with the technological innovations leading to interactive space(s) in the private and public domain.

[7] www.umapper.com/pages/geodart
[8] www.umapper.com/blog/?p=1596

ACKNOWLEDGEMENT(S) Again, thanks to the students Creative Technology, for (continually) providing the inspiration and motivation, and ultimately the workforce, for the CTSG.

REFERENCES

Bolter J.D and Grusin R. (2000), *Remediation – Understanding New Media*, MIT Press

Eliens A. and S.V. Bhikharie (2006), *Game @ VU – developing a masterclass for high-school students using the Half-life 2 SDK*, In: *Proc. GAME'ON-NA'2006*, Monterey, USA

Eliens A. Feldberg F., Konijn E., Compter E. (2007a) , VU @ Second Life – creating a (virtual) community of learners, In *Proc. EUROMEDIA 2007*, Delft, Netherlands

Eliens A., van de Watering M., Huurdeman H., Bhikharie S.V., Lemmers H., Vellinga P. (2007b), Clima Futura @ VU – communicating (unconvenient) science, In Proc. GAME-ON 07, Bologna, Italy

Eliens A., Huurdeman H., van de Watering M., Bhikharie S.V. (2008), XIMPEL Interactive Video – between narrative(s) and game play, In Proc. GAME-ON 2008, Valencia, Spain

Eliens A. and Ruttkay Z. (2008), Record, Replay & Reflect – a framework for understanding (serious) game play, In Proc. Euromedia 09, Brugge, Belgium

Eliens A. and Ruttkay Z. (2009), Math game(s) – an alternative (approach) to teaching math?, In Proc. GAME-ON 2009, Dusseldorf, Germany

Eliens A. (2010), Creative Technology – the CTSG: game design in 7 steps, submitted to GAME-ON Asia 2010, Shanghai, China

Flanagan M. (2009), *Critical Play – Radical Game Design*, MIT Press

Wardrip-Fruin N. (2009), *Expressive Processing – Digital Fictions, Computer Games, and Software Studies*, MIT Press

Anton Eliens is professor creative technology / new media at the University of Twente and coordinator of multimedia @ VU University Amsterdam. He has experience in web-based media, interactive video, and the development of serious games.

FULL LIFE-CYCLE AUTOMATIC ANIMATION GENERATION OF CHINESE TRADITIONAL ARCHITECTURE

Songmao Zhang
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
Beijing 100190, China
E-mail: smzhang@math.ac.cn

Kai Sun
Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology
Beijing University of Technology
Beijing 100022, China
E-mail: china_bjut_sunkai@hotmail.com

KEYWORDS
AI-based animation, Chinese traditional architecture, knowledge representation and reasoning, qualitative and quantitative animation description, full life-cycle automatic animation generation.

ABSTRACT

We propose the technology of full life-cycle automatic animation generation of Chinese traditional architecture and summarize a software system that implements this technology. Starting from a pseudo-natural language description of the architecture, until the generation of a 3D animation demonstrating the construction of this building, the whole process is performed automatically step by step, including; reasoning of the construction sequence, computation of the size and position of each piece of wood, generation of a qualitative animation specification, and generation of a quantitative animation specification. As a result, the software system realizes more than 180 architectures in three main types of Chinese traditional timber structures, thirteen types of Dougong, and six types of Chinese pagodas. In addition to standard architectures, our technology can, also be applied to generate animations for real-world buildings. Potential applications of such technology include demonstration systems in museums where the visitor describes what kind of building he wants to see and the system can accordingly construct the building in real time. We use artificial intelligence techniques including knowledge representation and reasoning for the Semantic Web to support the automation.

INTRODUCTION

AI-based animation studies how to apply artificial intelligence techniques to augment the automation and flexibility of computer animation generation. A lot of work has been presented since the early 90's, covering the (semi-) automatic generation of various elements of animation, including articulation and motion (Garcia-Rojas et al. 2006); human facial expressions (Pelachaud and Prevost 1994); camera control (He et al. 1996), dynamic environments (Gutierrez et al. 2005); video sequences (Arens and Magel 2003), human dialogues (Elliott et al. 2008), story and plot (Wang et al. 2006), background music (Nakamura et al. 1994), and Chinese Calligraphy (Xu et al. 2007). Many work deal with situations involving natural language, for example, using natural language instructions to manipulate virtual agents (Cassell et al. 2001; Wang and Panne 2006), text-to-scene conversion (Coyne and Sproat 2001; Johansson et al. 2005), and generating animations from natural language stories (Noma et al. 1992).

In this paper, we present a technology of full life-cycle automatic animation generation of Chinese traditional architecture. We have implemented a software system, which receives a pseudo-natural language description of an architecture and finally generates a 3D animation illustrating the construction of the building. This process features automatic reasoning and computation of each piece of wood and joinery among them. We demonstrate the feasibility of the automatic animation technology through various types of Chinese traditional architectures, including timber structures and pagodas. Potential applications of the technology and its merits and drawbacks are also discussed.

BACKGROUND

The study utilizes full life-cycle automatic animation generation technology from natural language stories (Lu et al. 2006) in the domain of Chinese ancient architecture.

Full Life-Cycle Automatic Generation of Animation

Ruqian Lu first proposed the full life-cycle computer-aided automatic animation generation technology in the late 80's, and his team implemented a system to demonstrate the technology in the mid 90's (Lu and Zhang 2002). Starting from a story written in some limited Chinese natural language, a system called SWAN covered the complete automatic process consisting of natural language processing, story understanding, animation plot design, character design, action planning, camera control and finally, generation of a 3D animation. Three languages were provided, so users could interact with the system at different levels of the transformation. The first is a limited Chinese natural language where the user can describe or change the story, the second is a qualitative animation representation language where the user can select the animation effects, and the third a quantitative animation representation language where the user can control the concrete parameters of the animation. A 10-minute animation cartoon was generated semi-automatically by SWAN within two weeks of time and was broadcast on China Central TV Channel in 1992. Full life-cycle animation generation involves many challenging AI tasks such as commonsense reasoning, story understanding, and automatic planning of articulation and movement of characters.

Chinese Traditional Architecture

Chinese traditional architecture is unique in the world and the construction of the timber structures is the core of the building. In general, a timber structure consists of three primary levels, the frame network of columns, the Dougong level, and the roof, where Dougong is a "unique structural element of interlocking wooden brackets" (http://en.wikipedia.org/wiki/Dougong). Dougong was originally devised to provide support for the roof in the Tang and Song Dynasty, and later has gradually become ornamental in the Qing Dynasty. Wooden pieces of Dougong, as well as those at other levels of the architecture, connect together by joints alone to form a stable structure. Ancient Chinese central governments actually released treatises to regulate the ranks of architecture and materials used. For example, a Wudian timber structure is of the highest rank and only royal buildings were permitted to adopt the Wudian style. The Taihe Temple in the Forbidden City is a typical complex Wudian building with a double-layered roof.

METHODS

We present every step in the whole life-cycle of automatic animation generation of Chinese traditional architecture as follows.

Information Extraction of User's Description

The user input allows a Chinese natural language-like description of the architecture, and we use information extraction techniques to obtain necessary data from the text. An example of the input can be "I want to see a grand Wudian style building, with five sections from left to right, and seven purlins from front to rear. The building has a surrounding veranda, and for Dougong I want Wucai type." Sections from left to right, represent the space between two columns, and the bigger their number the wider the building. Purlins are members of the roof and the bigger their number the taller the building. A veranda is a roofed open porch that surrounds the exterior walls and is optional for the Wudian style building. Wucai is one type of Dougong in the Qing Dynasty. The information extraction module of the system checks for integrity of the input data. For example, for all timber structures, both the number of left-to-right sections and the number of front-to-rear purlins have to be odd, and for pagodas, the number of layers is generally odd and the number of sides even.

Reasoning of Components and their Construction Sequence

A knowledge base, was built to represent the components and their relationships in Chinese traditional architecture. We used the Semantic Web techniques (http://www.w3.org/2001/sw/) (Berners-Lee et al. 2001) for knowledge representation and reasoning. More precisely, OWL DL (W3C OWL Working Group 2009) is used to specify an ontology (Brian 2006; Zhang and Bodenreider 2007) about Chinese traditional architecture, including subclass relations such as

Wudian -isa-> TimberStructure

and partitive relations such as

InnerColumn -part of-> WuDian

where *Wudian*, *InnerColumn*, and *TimberStructure* are all classes in the ontology. An OWL DL reasoner such as Pellet (http://clarkparsia.com/pellet) was used to ensure the logical consistency of the ontology. In addition to the ontology, which represents the conceptualization of the domain, knowledge about, for example, how many inner columns are needed in a Wudian building of *n* sections, *m* purlins and a veranda, is represented by the rule language SWRL (Horrocks et al. 2004) and we use Jess (Friedmann-Hill 2008) to do the rule reasoning.

To demonstrate the construction process we take into account every single piece of wood installed in the timber structure. Based on the data extracted from the user's description of the architecture, the knowledge base system infers what types of wooden components are needed and their respective numbers, and thus generates individuals for corresponding classes in the ontology. For example, for a Wudian building with five sections, seven purlins and a veranda, 25 types of wooden components are needed and a total of 5,437 pieces of wood are generated, including 14 individuals of inner column class and 22 individuals of outer column class. In order to specify every single piece of wood, we dissect the timber frame structure in three directions, vertical, horizontal from left to right, and horizontal from front to rear. Further we represent one piece of wood through a 4-tuple, i.e., the type of wooden component the wood belongs to, the start and end layer it comes across vertically, left-to-right horizontally, and front-to-rear horizontally, respectively. The 4-tuple representations of pieces of wood are important for deciding their construction sequence and their 3D sizes, positions and rotations.

Constructing Chinese traditional architecture has to follow a certain sequence so that every two pieces of wood that meet together can be jointed correctly. We use SWRL rules to specify the construction sequence of components, for example, inner columns must be installed before the outer columns.

Computation of the Size and Position in a 3D Scene

The animation we generate demonstrates the construction of a building in a 3D scene where pieces of wood fall one-by-one in certain sequence into their final positions in the architecture. We studied the architectural treatises of the ancient Chinese central governments (Guo 1998; Needham 1986) and summarized rules to compute the size of every type of wooden component, i.e., the three size values at the vertical, left-to-right horizontal and front-to-rear horizontal direction. The shape of wooden components can be rectangular, cylindrical or irregular, and for the irregular we compute the size of its bounding rectangle. The size of some component depends on the others, e.g., the vertical size of the inner column is the vertical size of the outer column plus some adjustment. For the purpose of joint connection, additional length is needed for some wooden components. For example, a purlin at the left-to-right direction is expected to cross-connect with a purlin at the front-to-rear direction, so the lengths of two purlins should include what is needed for their joint. Different timber structures have their own set of size computation rules.

Once the size is obtained, the position of every single piece of wood can be computed, i.e., the three axis values of the center of the wood in a 3D scene and the three rotation values. Since a building is constructed from bottom to top, we follow the corresponding sequence to compute the position. One of the most complex position computations lies in that for rafters. Firstly, rafters are installed side-by-side in parallel on the top of the roof, and only those positioned at the dissected points can be represented by our 4-tuple specification, called key rafters. Rafters between two key rafters, rather than in the first step for component generation, are not generated until now so that they have the same size as the key rafter, and their number and positions are computed based on the size and position of the key rafters. Secondly, rafters at every one of the four corners of the building have to be displayed to fit into an area of triangle, so these rafters differ not only in position but also in size. They can not be generated or computed until the size of the area of triangle is known. Figure 1 illustrates the rafters of a Wudian building generated by our system.

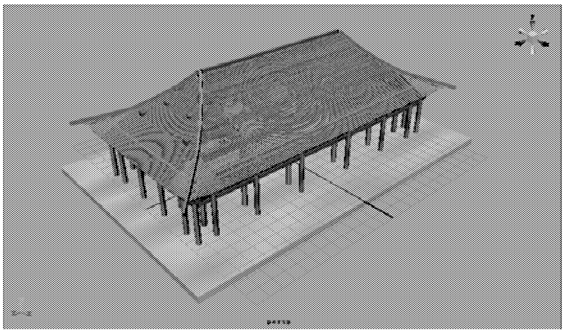


Figure 1: Computation of Rafters

Generation of the Qualitative Specification of Animation

Once the size and position of every piece of wood is obtained, we turn to generate the qualitative specification of the construction animation. In order to describe how a piece of wood moves in a 3D scene, its qualitative representation may look like:

```
moveTo(innerColumn1, fromPosition(above),
toPosition(<10, 25,10>), downwards, veryFast, t1,t2)
```

where the start position, direction, speed, and the start and end frame in the animation are all qualitative. If the user is not satisfied he or she can change the animation effect to, for example,

```
moveTo(innerColumn1, fromPosition(leftTop),
toPosition(<10, 25,10>), downwards, slowly, t1,t2)
```

For Dougong, the qualitative animation representation specifies the components and their qualitative locations and directions. Figure 2 shows examples of Dougong generated by our system, where in the upper left is a Dougong of the Song Dynasty, in the upper right of the Qing Dynasty, and at the bottom a complete Dougong level. A part of the qualitative description of Dougong can be as follows.

```
Component=Dadou, direction=horizontal, connected to=
        Component=Zhenxinguagong
        located on top of Dadou
        direction=vertical
        Component=Danqiao
        located on top of Dadou
        direction=horizontal
```

It specifies that the very bottom component of the Dougong is called Dadou and placed horizontally. On top of it there are two components, Zhenxinguagong and Danqiao, placed vertically and horizontally, respectively.

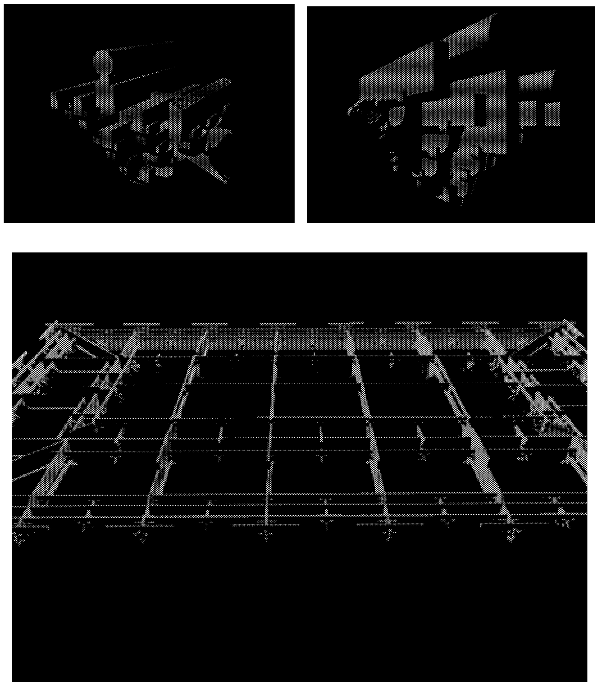


Figure 2: Animation of Dougong

Generation of Joints in a 3D Scene

Tenons and mortises are widely used in ancient China to connect two or more pieces of wood together so as to form stable timber structures. Various joints were devised, for connecting different types of wooden components at differing positions. We implemented ten types of joints and summarized rules for specifying where they are used and the computation of tenons and mortises in respective wooden components, as illustrated in Figure 3. What features is that the carpentering of the tenons and mortises is performed, rather than beforehand, on the fly in real time when the final animation is shown through the Boolean operations in the Maya animation software. This saves us from having to generate in advance the animation object files for all possible pieces of wood with all possible kinds of tenons and mortises.

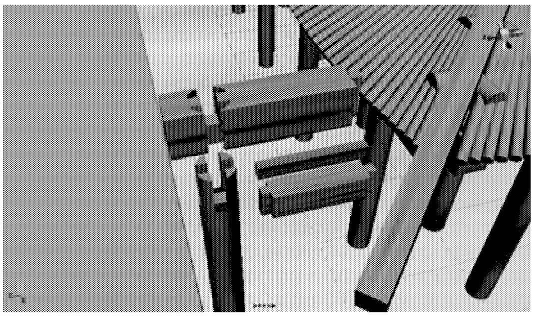


Figure 3: Generation of Joints

Generation of the Quantitative Specification of Animation

From the qualitative description of animation, concrete 3D values can be computed for every component, including the size, position and rotation. For the inner column described above, quantitative computation presents that from frame 1 to frame 15 it moves evenly from position <10, 45,10> to <10,25,10> without any rotation or scaling. For the Dougong described above, a part of its quantitative description is given as follows.

```
Component=Dadou
        modelName="dadou.mb"
        Size=<30,20,30>, Transfer=<0,0,0>,
        Scale=<1,1,1>, Rotate=<0,0,0>
Component=Zhengxinguagong
        modelName="zhengxinguagong.mb"
        Size=<62,20,12.4>, Transfer=<0,12,0>,
        Scale=<1,1,1>, Rotate=<0,270,0>
```

We have built a database for storing model objects in animation and their size data, e.g., "dadou.mb" and "zhengxinguagong.mb" are model object files in Maya. For regular wooden components, only one rectangular object and one cylindrical object are stored in the database, and they are scaled up to match the computed size for every specific individual piece of wood whenever needed.

Generation of Animation Files

The quantitative animation specification is converted into the ".ma" scene file format of the Maya animation software, which is ready for rendering. Figure 4 shows the animation of a Wudian building. The quantitative representation basically contains all the data needed for animation, and thus it can be transformed into any textual animation file format. For example, in order to provide fast real-time preview of the animation, we also convert the quantitative representation into VRML. Specifically, for joints among pieces of wood, script files ".mel" of Maya are generated to implement the Boolean operations in order to form the 3D tenons and mortises on the fly.

Figure 4: Animation of a Wudian Timber Structure

RESULTS

Generating Animations of Standard Buildings

Three of the main timber structure types of Chinese traditional buildings were realized by our system, two of high ranks with the Dougong level and one of low rank without Dougong. From the user's description of the number of sections from left to right and the number of purlins from front to rear, together with other data, the system generates an animation to show how and in what sequence every piece of wood is installed to finally complete the whole timber structure.

In addition to the timber structures, thirteen types of Dougong can also be automatically constructed in our system, including seven types of the Song Dynasty and six of the Qing Dynasty.

Moreover, six types of Chinese pagodas can be built in the animations generated by our system. The user describes the number of layers of the pagoda, its shape being a quadrangle, hexagon, or octagon, the type and shape of the top of the pagoda, and if some decorations at the top. The system accordingly computes and builds the pagoda layer-by-layer from bottom to top in animation. Figure 5 shows examples of the animations of pagodas.

Our system currently covers more than 180 types of Chinese traditional architecture.

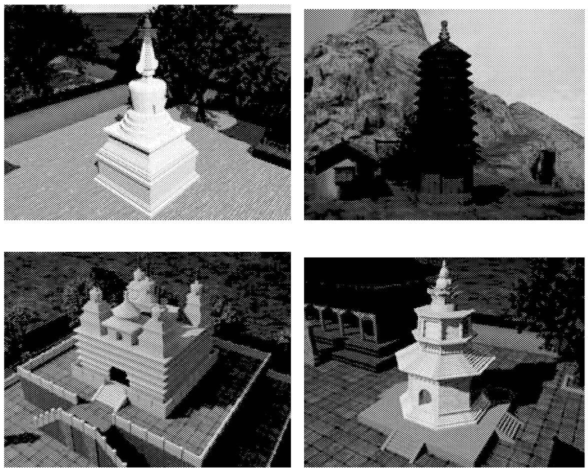


Figure 5: Animation of Chinese Pagodas

Generating Animations of Real-World Buildings

All the above-mentioned architectures are standard buildings specified in the architectural treatises released by the ancient Chinese central governments. However, the construction of the real-world buildings has never exactly followed the standard regulations. Deviations could not be avoided, due to the limitation of materials available and sometimes were intended by the innovations of carpenters. We selected the Foguang Temple in Shanxi Province in China (http://en.wikipedia.org/wiki/Foguang_Temple) and showed how our technology can be applied to generate animations for real-world traditional buildings. The East Hall in Foguang Temple built in 857 AD during the Tang Dynasty is the third earliest purely wooden building in existence in China. With the measures provided by Tsinghua University, we manually transformed the data of each piece of wood into our own 4-triple representation and the size and position representation in a 3D scene. The sequence of the construction is reasoned by our system. From this step on, the system performs in the same way as for the standard buildings, i.e., to generate a qualitative and then quantitative animation description, and lastly the animation software scene file that can be rendered to obtain the final video, as shown in Figure 6.

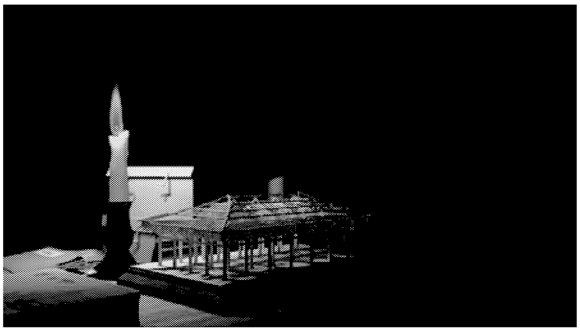


Figure 6: Animation of the East Hall of the Foguang Temple

DISCUSSION AND CONCLUSIONS

We present the full life-cycle of automation from a textual description of the structure of a Chinese traditional architecture and finally into a 3D animation of the construction of the building. The steps include information extraction of the user's pseudo-natural language description, reasoning of the components and the construction sequence of these components, generation of joints among components, computation of the size, position and direction in a 3D scene, generation of a qualitative animation specification, generation of a quantitative animation specification, and generation of scene file formats of animation softwares ready for rendering or real time preview. All these automatic transformations step-by-step actually present a semantic understanding process for the domain of Chinese traditional architecture supported by the knowledge representation and reasoning techniques. The Semantic Web ontology and rule formalisms were used to represent a knowledge base ready for the future sharing, re-use and publication. We demonstrated the feasibility of the technology in three main types of Chinese traditional timber structures, thirteen types of Dougong, and six types of Chinese pagodas. Moreover, our technology can also be applied, to generate animations for real-world buildings.

Potential applications of our automatic animation technology include demonstration systems in museums where the visitor can describe what kind of building he or she wants to see and the system accordingly constructs the building in real time. In case the user is not satisfied with the animation, he or she can change the input and the system would immediately reflect the revision and generate a different building. Furthermore, the modification can be performed not only at the pseudo-natural language interface but also in other steps of transformation, e.g., the sequence of construction can be changed in the qualitative representation.

Compared with other AI-based animation systems and tools, our technology has the feature of scalability. To expand the system to generate animations for another type of architecture, one only needs to summarize the rules for sequence reasoning and the rules of size and position computation, while all the other modules in the system can be used directly without updates. Furthermore, if targeting another animation software, one only needs to code the transformation from our quantitative representation into the textual file format of that animation software. On the other hand, the disadvantages of our system include that for real-world buildings, we have to manually transform the measuring data of the components into our representations. Methods for automatically extracting components from the point cloud data of a structure (Zhu et al. 2007) could be complementary to our technology. Moreover, we would add the automatic camera control and light specification generation to the animation in our future work. The most challenging task would be the automatic planning of human articulated motions in the construction animation where quite often a group of workers have to cooperate with each other to accomplish one complex construction action, for example, the erection of columns.

ACKNOWLEDGEMENTS

This project was supported by the National Key Technology R&D Program of China, Knowledge Innovation Program of the Chinese Academy of Sciences, MADIS of the Chinese Academy of Sciences, and the Key Laboratory of Multimedia and Intelligent Software Technology at Beijing University of Technology.

We sincerely thank Prof. Ruqian Lu from the Chinese Academy of Sciences, Prof. Chunnian Liu from the Beijing University of Technology, and all of the graduate students who participated in the design and development of the system, including Gongjin Wei, Meifang Yin, Weifeng Wang, Liang Kong, Jia Sun, Lihuang Gong, Yan Wang, Jia Feng, Shebiao Liu, Tianzhu Liang, Weijing Bai, Bo Gu, Rongrong Sun, and Shan Zhu.

REFERENCES

Arens, M. and H. Magel. “Behavioral Knowledge Representation for the Understanding and Creation of Video Sequences”. In *Proceedings of KI 2003*, LNAI 2821, 2003, 149-163

Brian, E. “Constructing an Ontology”. *Topics on General and Formal Ontology*. 2006, Polimetrica Publisher, Italy.

Berners-Lee, T.; J. Hendler; and O. Lassila. “The Semantic Web”. *Scientific American*, 284(5), 2001, 34-43

Cassell, J.; H. Vilhjalmsson; and T. Bickmore. “BEAT: the Behavior Expression Animation Toolkit”. In P*roceedings of SIGGRAPH 2001*, Los Angeles, 2001, 477-486.

Coyne, B. and R. Sproat. “WordEye: An Automatic Text-to-Scene Conversion System”. In *Proceedings of SIGGRAPH 2001*, Los Angeles, 487-496

Elliott R.; J.R.W.Glauert; and J.R.Jennaway. “Linguistic Modeling and Language-Processing Technologies for Avatar-Based Sign Language Presentation”. *Univ Access Inf Soc* (2008) 6:375-391

Friedman-Hill, E. *Jess: the Rule Engine for the Java Platform*. Jess Manual, 2008, http://www.jessrules.com/jess/docs/71/

Garcia-Rojas, A.; F. Vexo; D. Thalmann; A. Raouzaiou; K. Karpouzis; and S. Kollias. “Emotional Body Expression Parameters In Virtual Human Ontology”. In *Proceedings of the 1st International Workshop on Shapes and Semantics*, 2006

Guo, Q. “Yingzao Fashi: Twelfth-Century Chinese Building Manual”. *Architectural History: Journal of the Society of Architectural Historians of Great Britain* (Volume 41 1998): 1–13

Gutierrez, M.; F. Vexo; and D. Thalmann. “Semantics-based Representation of Virtual Environments”. *International Journal of Computer Applications in Technology*, 2005, Vol.23, Nos. 2/3/4, 229-238

He, L.; M. Cohen; and D. Salesin. “The Virtual Cinematographer: A Paradigm for Automatic Real-Time Camera Control and Directing”. Technical Report, August 1996

Horrocks, I.; P.F.Pater-Schneider; H.Boley; B.Grosof; and M.Dean. *SWRL: A Semantic Web Rule Language Combining OWL and RuleML*. W3C Member Submission, 2004, http://www.w3.org/Submission/SWRL/

Johansson, R.; A. Berglund; M. Danielsson; and P. Nugues. “Automatic Text-to-Scene Conversion in the Traffic Accident Domain”. In *Proceedings of 2005 International Joint Conferences on Artificial Intelligence (IJCAI-05)*, 2005, 1073-1078

Lu, R. and S. Zhang. 2002. *Automatic Generation of Computer Animation*. LNAI 2160, Subseries of LNCS, Springer.

Lu, R.; D. Zeng; and F. Wang. “AI Research in China: 50 Years down the Road.” *IEEE Intelligent Systems*, May/June 2006, 91-93

Nakamura, J.; T. Kaku; K. Hyun; T. Noma; and S. Yoshida. “Automatic Background Music Generation based on Actors’ Mood and Motions”. *The Journal of Visualization and Computer Animation*, 1994, 5, 247-264

Needham, J. 1986. *Science and Civilization in China: Volume 4, Part 3*. Taipei: Caves Books, Ltd.

Noma, N.; K. Kai; J. Nakamura; and N. Okada. “Translating from Natural Language Story to Computer Animation”. In *Proceedings of SPICIS’92*，1992，475-480

Pelachaud, C. and S. Prevost. “Sight and Sound: Generating Facial Expressions and Spoken Intonation from Context”. In *Proceedings of the 2nd ESCA/AAAI/IEEE Workshop on Speech Synthesis*, New York, USA, September, 1994

W3C OWL Working Group. O*WL 2 Web Ontology Language Document Overview*. W3C Recommendation 2009, http://www.w3.org/TR/owl2-overview/

Wang, H.; C. Cao; and P. Yu. “Representing and Using Character Feature Rules in Automatic Story Generation”. In *The 5th WSEAS International Conference on Applied Computer Science*, 2006

Wang, Z. and M. Panne. “Walk to Here: A Voice Driven Animation System”. In *Proceedings of ACM SIGGRAPH Symposium on Computer Animation 2006*.

Xu, S.; H. Jiang; F. C.M. Lau; and Y.Pan. “An Intelligent System for Chinese Calligraphy”. In *Proceedings of AAAI 2007*, 1578-1583.

Zhang, S. and O. Bodenreider. “Experience in Aligning Anatomical Ontologies”. *International Journal on Semantic Web and Information Systems*, 3(2), 2007, 1-26.

Zhu, L.; R. Shi; and K. Zhou. “Rule-based 3D Modeling for Chinese Traditional Architecture”. In *Proceedings of the 2nd ISPRS International Workshop on 3D Virtual Reconstruction and Visualization of Complex Architectures* (3D-ARCH 2007), 2007

AUTHOR BIOGRAPHY

SONGMAO ZHANG got her PhD degree in 1992 at the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and she is now a full professor at this academy. She was a visiting scholar in universities and institutions in the US, Australia and Europe. Her research interests mainly include ontology and the Semantic Web, AI-based automatic animation, data mining, and natural language understanding.

KAI SUN got his undergraduate degree in 2007 at Beijing University of Technology, and he is now a graduate student at this university. His research interests mainly include AI-based automatic animation, natural language processing, and machine learning.

Priority-Based Level of Detail Approach for Animation Interpolation of Articulated Objects

Antoine Abi Chakra
Xin Li

Computer Science Department
DigiPen Institute of Technology
Redmond, Washington, U.S.A.

KEY WORDS

Animation, Distance-Based, Level of Details (LOD), Articulated Character, Real-Time Simulation, Video Games

ABSTRACT

This paper presents an LOD technique applicable to articulated characters in real-time graphical simulation or video game software. This new method does not eliminate the polygon counts of the graphical model due to the change of distance to the viewer. Rather, the rendering expense is cut by the reduction of the animation interpolations between key frames based on a priority number that is calculated and assigned by taking several factors into account, including rotation, translation, scaling, number of children and depth in the hierarchical structure. An analysis to those sub-components of the hierarchical animation is conducted and an equation is established to qualify the contribution to visual effects of such animation sequence. The benchmark tests demonstrate that, under certain conditions, our approach can significantly reduce the number of interpolated with no noticeable visual artifacts.

1. INTRODUCTION

Since the introduction of 3D graphics, special visual effects have flooded the video game and graphical application. The number and types of such visual effects that can be employed in any particular game, such as particle systems, HDR, anti-aliasing, real-time shading, physically-based lighting, natural pronominal simulation, uneven surface mapping, etc., is highly dependent on the hardware configuration of the computer, including the CPU frequency, the GPU processing power, on-board memory size and bandwidth, the programmable shader units. On the other hand, the software efficiency and optimization are also the major factors contributing to the rendering capability.

With the transition from 2D to 3D, computer hardware was (and still is) being pushed more and more to produce better picture quality. Each generation of graphic cards adds new features to its predecessors, new techniques are easier or faster to implement, and the goal remains to same: Squeezing more and more power juice into 3D applications. If programmers relied only on hardware power, then we would be forced to wait for the next generation of graphic cards to get faster frame rates and better looking games. Fortunately, that is not the case since there are many different software algorithms well researched and developed to alleviate hardware processing. One of such algorithms is the so-called “Level of Detail” (LOD) approach. Unlike other optimization techniques, the LOD method deals with visible polygons. After performing all other “polygon reduction” algorithms, it takes the distance that separates a visible model from the camera into consideration before rendering it. The greater the distance is, the fewer polygons of that object (such the less details) is rendered [1], [2] and [8].

A great deal of research has been devoted to this area. In this paper, we shortly mention some popular LOD methods previously done to reduce the polygon counts. Then we present in detail a new LOD technique designed to optimize the rendering process for articulated objects with the animated motions. Unlike any traditional LOD methods, our approach does not reduce the polygon counts of the object based on the distance to the viewer. Rather, the rendering time is saved by the reduction of the animation interpolations between key frames. The method is called the priority-based LOD approach since the decision of the interpolation reduction is made based on a priority number that is calculated and assigned by taking several factors into account, such as rotation, translation, scaling factor, number of children and depth in the hierarchical structure.

After the introduction section, we briefly review the hierarchical structure that is popularly used to represent articulated objects in video game software in Section 3. The degree of contribution that different transformation parameters make to the overall shape and visual effects of such articulated model during an animation sequence is examined in Section 4, which leads to the discussion on the analytic model to qualify such contribution in Section 5. In the sequential sections, we also present a method integrating this analytic model with the LOD distance as well as some implementation considerations. The result of benchmark tests and possible future research are also discussed in the paper.

2. PREVIOUS WORK

There are three main types LOD techniques.

2.1 Discrete LOD

Discrete LOD reduces the number of polygons of 3D models in a discontinuous way. The artist is responsible of creating several models for all objects, each with a different number of polygons. At run time, these models are switched depending on the distance separating them from the camera [10]. Some techniques are developed to avoid the popup effect when switching models, including late switching, hysteresis, alpha blending, etc. [4].

2.2. Continuous LOD

Continuous LOD deals with the popping problem which occurs in the discrete LOD. Instead of having several static meshes, the method uses the original mesh only and decreases object’s polygon count gradually as it gets smaller on the screen with simplification operators such as edge collapse and vertex-pair collapse [3], [5], [6], [7] and [9].

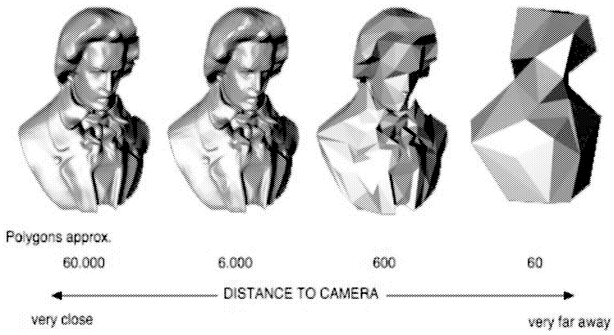


Figure 1: Discrete LOD models

2.3 Non geometric LOD

Other non geometric-based LOD techniques can are also employed to save the rendering time by reducing the number of effects as the objects becomes smaller on the screen, such as utilizing 2D imposters to replace the 3D objects, etc

3. MOTION LOD

3.1 Hierarchy structure

Articulated models are widely used in applications and video game and other 3D graphical applications. In the articulated modeling technique, the 3D object is represented and stored as a tree-like structure. The root of the tree can have any number of children bones, which in turn are linked with a number of children bones of their own.

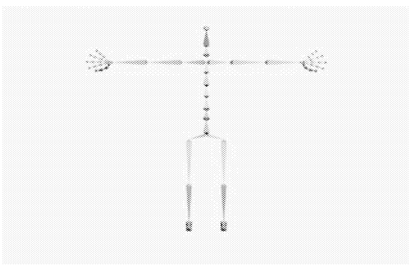


Figure 2. Hierarchical structure of an articulated object

One technique to represent the motion of such articulated object is to associate each bone or joint of the hierarchical structure with a unique set of three dimensional transformations referred as key frames, along with a respective time stamp for each key frame. At run time, a set of transformation data (combining scale, rotation and translation) is created by interpolating between 2 key frames at each time_stamp. This transformation data is often generated (or converted) into the form of matrix and concatenated with that of its children's along the hierarchical structure. Obviously, any change in the transformation of any children bone will be propagated to all the bones located beneath that bone in the animation tree.

Obviously, the discrete and continuous LOD methods introduced in the previous sections can be applied individually to each part of the articulated model. However, the intention of our work is to enhance the optimization of the model by targeting another "expensive point" in the process, which is the transformation interpolation. This can be done by replacing the entire object by another one containing fewer bones when it becomes smaller on the screen, thus less transformation interpolations at each frame.

4. PRIORITY-BASED LOD

As explained previously, each bone in the hierarchy computes its own transformation matrix by interpolating between key frames, including scale, rotation and translation parameters. In general, the number of interpolations depends on the number of time "ticks" between the pair of key frame animations at the run-time. This number, however, can be reduced, without any noticeable visual anomaly if handled properly, as the rendered image of the model on the screen becomes smaller.

A simple method to achieve this goal is to reduce the number of interpolations for all the bones simultaneously as a hierarchical object moves away from the camera. The drawback of such naïve approach is that, since not all the bones of the object contribute to the change of animation equally, the possibility of visual anomalies introduced by this method could forbid further reduction of interpolations at the bones that contribute less to the motion behavior.

To determine the degree of contribution to the overall shape of the articulated model during an animation, we conduct an analysis in main factors such as rotation, translation, scaling, and other "states", including the depth in the hierarchical structure and the number of children. We first examine how each factor individually affects the animation interpolation from one key frame to the next, combine all factors with different weights to form a so-called priority equation, then utilize the equation to implement a priority-based algorithm that determines the number of the interpolation frame reduction.

4.1 Rotation

In the key frame based animation, each bone has its independent rotation parameters. Assume that a bone has a rotation (R_A) of 90 degrees in 2 frames while another bone has a rotation (R_B) of 40 degrees in the same number of frames. (See Figure 3 below.) Obviously, skipping a certain number of interpolation steps from rotation (R_B) would have less visual effect than skipping the same number of interpolation for rotation (R_A), simply because the fact that the degree of transformation of rotation R_A is much greater and therefore noticeable than the one of rotation R_B. Thus, our animation level of detail algorithm takes the rotation degree difference between consecutive key frames into consideration when determining how much a certain bone contributes to the overall shape of the animated model.

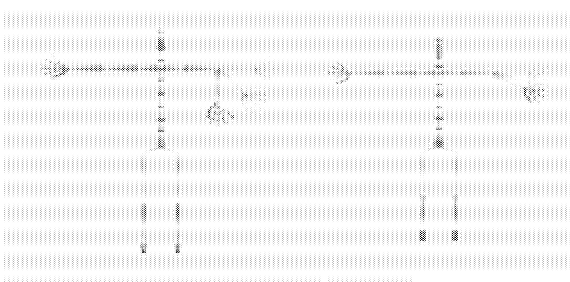


Figure 3: Difference between bone rotations

4.2 Translation

In an animated model, each bone has its own set of translation vectors per key frame, which is independent from other bones' translation vectors. We should check the translation effect on the overall shape of the articulated model similarly to the way we checked the effect of rotating a bone has on it. Assume that a bone is translated by a vector T_A=<20,0,0> in 2 frames while another bone is translated by a vector T_B=<5,0,0> during the same number of frames. (See Figure 4). Obviously the greater the translation vector is, the more it is visible when interpolating the key frames. Therefore, skipping a certain number of interpolation steps for bone affected by the translation vector T_B won't be as noticeable as skipping the same number of transformation interpolations from the bone affected by the translation vector T_A. Consequently, the translation transformation difference between 2 consecutive key frames is taken into consideration when determining the degree of contribution of each bone to the shape of the animated model.

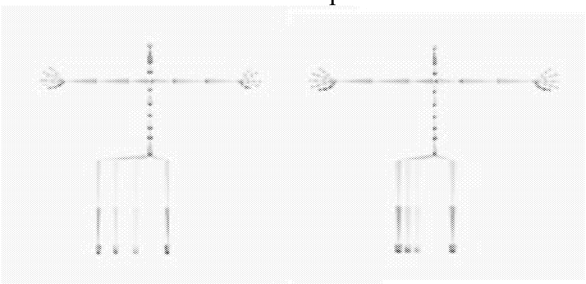


Figure 4: Difference between bone translations

4.3 Scaling

In the hierarchical model, each bone has a set of scaling values as a component of transformations. Since scaling a bone directly affects the offset of all the vertices associated with it relatively to the center, the difference between the scaling values of 2 consecutive key frames has a direct impact on the overall shape of the articulated model. Let's assume that one bone is to be transformed by the scaling factor S_A=<2.5, 2.5, 1> during 2 frames while another bone is to be transformed by a vector S_B=<1.5, 1.5, 1> during the same number of frames. Obviously skipping the same number of interpolated frames for both scaling transformations S_A and S_B will have a different visual effect on the animation model.

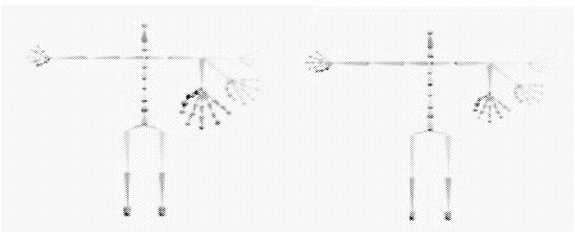


Figure 5: Difference between bone scaling

In conclusion, the scaling factor difference between 2 key frames should also be taken into consideration when determining the contribution of each bone to the shape of the animated model.

4.4. Number of children

In the articulated model description, we mentioned that the bones have a set of children bones. Each one of these children bones will concatenate its parent bone's transformation data with its own, in order to generate its final transformation data. This final transformation data of the bone in question will be passed to its own child nodes that will do the same process, until this propagation traverses all the bones of the articulated model.

It is clear that the transformation matrix of any bone B_d (d is the depth of the bone B), built by concatenating its translation, rotation and scaling transformations, will be propagated to all the bones located under the bone B_d. (See Figure 6 below.)

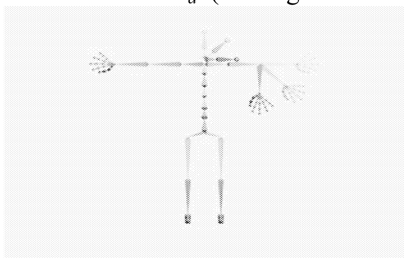


Figure 6: Difference between numbers of Children bones.

Because each bone's final transformation matrix needs its parent transformation matrix, then skipping a transformation interpolation for any bone would affect the final transformations of all the bones that are direct or indirect children of that bone. Thus we can conclude that skipping a transformation matrix interpolation of a bone having many direct and indirect children bones will have a greater negative visual effect on the overall shape of the articulated object.

The error term introduced by skipping interpolations in the hierarchical structure is relative to the total number of direct and indirect children bones and can be quantified as following:

B^{ω}

ω = Number of Direct and Indirect Children of bone B

$\phi_{Per\ Child}$ = Error factor per bone

$$\delta^{B^{\omega}}_{Number\ Of\ Children} = \omega \times \phi_{Per\ Child}$$

4.4 Depth in the hierarchy

The higher level a bone is in the hierarchy, the greater its transformation matrix influences the rest of the model. As explained previously, the animation model is saved as a tree structure, where each bone computes its final transformation matrix by concatenating its own local transformation matrix with its parent space transformation matrix. This means that any change in the transformation of a bone B_d (d is the depth of the bone B) will have a direct impact on all children bones of B_d. (See Figure 7 below).

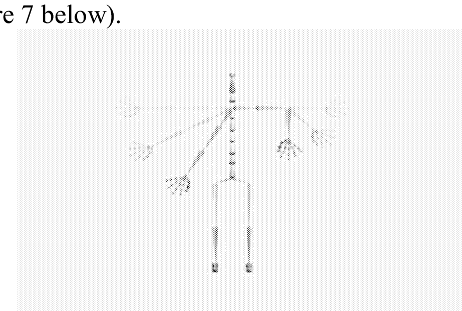


Figure 7: Difference associated with the depth

Having this in mind, it seems logical that skipping certain number of transformation interpolations of a bone located at a higher level in the animation hierarchy will distort the final shape of the animation model much more than skipping the same number of transformation interpolations of a bone located at the deeper level in the animation hierarchy. In other words, the error term introduced by skipping interpolations of any bone in the hierarchical structure is relative to the depth of the bone in the structure and can be quantified as following:

B_d

d = Depth of bone B

D = Maximum depth of the animation

ϕ_{Depth}= Error factor per single depth difference

$$\delta_{Depth} = \phi_{Depth} \times \frac{D}{d}$$

4.5 Combining the factors

As analyzed above, computing how much a bone contributes to the overall shape of the articulated model is done by checking how much skipping some of its transformation interpolations affect the results of the animated model. The greater this visual effect is, the more the bone is considered "*important*". Therefore, each bone should have certain *priority* based on its contribution when producing the final shape of the animation. This priority value determines how much a bone will "skip" its transformation interpolations. In other words, the higher the priority of a bone is, the more important it will contribute to the final shape of the animated object, therefore, the less frequent its transformation interpolation should be skipped.

Since each key frame of an animation of a hierarchical structure is independent, the current transformation depends solely on the transformations of its previous and next key frames. It implies that having a single priority for each bone does not seem to be enough. In other words, during an animated sequence, a bone can be transformed (i.e. change the orientation, size or location) in a great deal between certain key frames, while having a relatively minor transformation change between other key frames. Therefore, each key frame of each bone will have its own priority, whose value will alter from time to time during a simulation, depending on how much its transformation is "different" from the previous key frame's transformation.

5. PRIORITY PAMAMETERS

Based upon the above analysis, priority values are computed at run time to take into account the major factors that would contribute to the visual effect of the animation sequence once interpolated between pairs of key frames as listed below:

- Translation priority
- Rotation priority
- Scale priority
- Number of children priority
- Depth priority

This function is the governing equation when determining how and when the transformation interpolation should be skipped to save time once the animated object is distant to the camera. In practice, the priority value is also combined with a user-defined term to allow some flexibility and gain the control when the animated sequence is interpolated.

In this section the algorithm that implements this priority function is described in six steps as follows.

5.1 Initialization

In order to assign a priority, some information on the entire animation and structure of the model is gathered during the load time. The process starts from the root of the articulated object and loop through all the bones. For each bone, it scans through all animation key frames in order to determine the "maximum transformations" values, including the following parameters between 2 consecutive frames of any bone in the structure:

- maximum translation difference,
- maximum rotation,
- maximum scale,
- maximum number of direct and indirect children, and
- maximum depth value of the articulated model.

Once computed, those five values are stored for later use.

5.2 Translation

As mentioned previously, each key frame has its own translation vector. The length of this translation vector is computed in order to determine how much each bone is moving relatively to its parent bone. The greater the length of the translation vector, the more distant the bone is from parent. If the length of the translation vectors of the consecutive key frames is very different, then the bone is considered to be "agitated".

Let $\overrightarrow{T_n}$ be the translation vector of the bone during the current key frame, and $\overrightarrow{T_{n+1}}$ be the translation vector of the next key frame. $\overrightarrow{\Delta_T}$ is the difference between $\overrightarrow{T_{n+1}}$ and $\overrightarrow{T_n}$. Finally, l_T is the length of the vector $\overrightarrow{\Delta_T}$, which is computed as follows:

$$\overrightarrow{\Delta_T} = \overrightarrow{T_{n+1}} - \overrightarrow{T_n}$$

$$l_T = \left|\overrightarrow{\Delta_T}\right|$$

After computing the translation amount of the bone from key frame n to key frame n+1, the length of that translation difference is compared to the animation's current maximum translation difference and the result is saved in case for later use.

If (l_T > MaxTranslation) then MaxTranslation = l_T

5.3 Rotation

Besides a translation vector, each animation key frame has rotation parameter, usually saved in the form of a quaternion. In order to compute the difference between two rotations, the dot product of the two quaternions representing the rotation transformation of the key frame n and n+1 is computed. The greater the value of this dot product is, the more the bone is rotating around its parent. Therefore the rotation priority for this particular key frame is relative to that quaternion dot product. If the dot product of the quaternions of any two consecutive key frames is high, then again the bone is considered "agitated".

Let $\overrightarrow{R_n}$ be the rotation quaternion of the bone during the current key frame, and $\overrightarrow{R_{n+1}}$ be the rotation quaternion of the next key frame. β is the dot product result between $\overrightarrow{R_{n+1}}$ and $\overrightarrow{R_n}$

$$\beta = \overrightarrow{R_{n+1}} . \overrightarrow{R_n}$$

After computing the dot product of 2 quaternions, we compare the result to the animation's current maximum rotation dot product and replace it if it is greater.

if (β > Maximum Rotation) then Maximum Rotation = β

5.4 Scale

The scaling parameter is usually saved as a 3D vector. The greater the scale vector, the more the bone changes its geometry during the animation. If the magnitude of the scaling vectors of the consecutive key frames is very different, then the bone is considered to be very varying scale wise.

Let $\overrightarrow{S_n}$ be the translation vector of the bone during the current key frame, and $\overrightarrow{S_{n+1}}$ be the translation vector of the next key frame. $\overrightarrow{\Delta_S}$ is the difference between $\overrightarrow{S_{n+1}}$ and $\overrightarrow{S_n}$. Finally, l_S is the length of the vector $\overrightarrow{\Delta_S}$.

$$\overrightarrow{\Delta_S} = \overrightarrow{S_{n+1}} - \overrightarrow{S_n}$$
$$l_S = \left|\overrightarrow{S_T}\right|$$

After the scaling amount of the bone from key frame n to key frame n+1 is computed, the length of that scaling difference to the animation's current maximum scaling difference is compared and saved for later analysis.

If (l_S > Maximum Scale) then Maximum Scale = l_S

5.5 Maximum number of children

Since this value is only related to the hierarchical structure of the articulated model, the traversal through all the key frames of the animation sequence is not necessary. However, note that, when calculating the maximum number of children of a bone, all its direct and indirect children must be counted. This can be easily done by recursively traversing the tree structure, counting the number of the direct children, and adding maximum number of children of all direct children to the total count.

5.6 Depth

The depth of a bone in the hierarchical structure is the depth of the sub-tree with this bone as the root. In other words, it can be easily computed by counting the maximum length of all paths from this node to its deepest leaf node.

if (Current Depth > MaxDepth) then MaxDepth = Current Depth

5.7 Combining Priorities

Now that all the required information on the articulated model and its animation is gathered, we can compute the per key frame priority of each bone. The per-key-frame priority represents how much the key frame in question contributes to the overall shape of the animated character. The more a bone is agitated during a certain key frame, the higher that its priority will be, and the less chance the transformation interpolation will be skipped during the animation. On the other hand, a relatively low priority will allow the bone to skip a greater number of transformation interpolations to save time.

The priority is computed using the same criteria that were used to collect the articulated model's information, namely, translation, rotation, scaling, number of children, depth. A user-defined priority is also added to the final priority value. Six variables are used to represent the values as follows:

P_T = Translation Priority
P_R = Rotation Priority
P_S = Scale Priority
P_D = Depth Priority
P_{Num} = Number of Children Priority
P_{User} = User-Set Priority
$P_{b_i^j}$ = Priority of key frame j of bone j

In order to compute the value of each sub priority, the ratio between the key frame data and the maximum data that were collected in the previous step are calculated and saved. Bones that generate the highest criteria value will eventually end up with relatively high priorities, allowing them to minimize the number of skipped transformation interpolations.

Starting from the Root
For each children bone b_i
For each key frame b_i^j

$$P_T = \overrightarrow{\Delta_T^{b_i^j}} \text{ / MaxTranslation}$$
$$P_R = \beta^{b_i^j} \text{ / MaxRotation}$$
$$P_S = \overrightarrow{\Delta_S^{b_i^j}} \text{ / MaxScale}$$
$$P_{Num} = \text{NumOfChildren / MaxNurOfChildren}$$
$$P_D = \text{Current Depth / Max Depth}$$
$$P_{b_i^j} = (P_T + P_R + P_S + P_D + P_{Num} + P_{User}) / 6$$

5.8 Weighted Average

After computing the value of the six sub-priorities, the final priority of each key frame is obtained by averaging those sub priorities. One approach is to treat these sub priorities equally by explicitly assigning the same *weight* for each one, i.e. $w = 1/6$,

$$P_{b_j^i} = wP_T + wP_R + wP_S + wP_D + w\,P_{Num} + w\,P_{User} \quad 6$$

However, it can be also done differently when the user wants to emphasize on any particular sub priority. For example, if the key frames of the bones of the animated model are pre-known (with information provided by the artist) to be very different rotation wise, which means the difference between 2 consecutive rotations is generally very high compared to the translation or scaling difference, it would be more logical to assign the rotation sub priority with a greater.

The weighted average of the 6 sub priorities allows users to have a finer level of flexibility (compared to the user-set sub priority factor) when determining the final sub priority of each key frame of the animated model's bones. In practice, we define the following six weight factors with different values. The sum of all the weights is equal to 1.0

w_T = Weight of the Translation sub-priority
w_R = Weight of the Rotation sub-priority
w_S = Weight of the Scale sub-priority
w_{Num} = Weight of the number of children sub-priority
w_D = Weight of the Depth sub-priority
w_{User} = Weight of the User-set sub-priority

It's the user's responsibility to adjust the weights of each sub priority to allow the animation level of detail algorithm to adapt to different circumstances. For example, if the animation manipulates the scale factors more frequently with greater differences than it manipulates the translation factors, then the weight associated with the scale sub priority should obviously be greater than translation's weight factor.

6. APPLYING PRIORITY

6.1 Averaging priorities

The static priority of each key frame for bones is used to manipulate the LOD. Although the entire hierarchy's information is considered to compute the static priorities, the LOD of each bone is manipulated separately. As the first step, the entire tree is traversed and the following information is gathered:

- *MaxTranslation*: Maximum translation during one frame
- *MaxRotation*: Maximum rotation during one frame
- *MaxScale*: Maximum scaling during one frame
- *MaxNumberOfChildren*: Maximum number of children bones per bone
- *MaxDepth*: Maximum depth of the animated model

In step 2, the algorithm steps through all key frames of each bone to compute the sub-priorities described in Section 5.7. Finally, the sub priorities are averaged in order to determine the final static priority of each key frame of each bone.

6.2 Applying with Distance

The idea of any LOD technique is to reduce the details that can no longer be perceived. This occurs when the object's final projected size of the screen is relatively small compared to its original size. In a 3D scene, this occurs when the distance separating the object and the camera increases. Since object's locations in the 3D scene vary unpredictably in video games, the final level of detail is frame-based. In other words, it is totally independent from the previous and next frame's level of detail.

In order to use the object's position when computing its current frame's level of detail, we need to specify a distance range: [*MinDistance – MaxDistance*]. This LOD range is relative the viewer, which is computed for each object as **dynamic ratio Ω**. If the object's position is less than the lower bound of the distance range, then it is considered to be greatly visible. In this case, every bit of its detail is perceived. Therefore, any object closer to the viewer than the range's lower bound won't have its level decreased, which means all of its bones will interpolate their transformations normally as they do if there wasn't any level of detail technique applied. That is : **Ω** = 1.0

If the object's position is greater than the upper bound of the distance range, then it is considered to be barely visible, and all of its original detail is barely perceived, if even at all. Therefore, any object whose distance separating it from the viewer is greater than the upper bound of the LOD range will skip all of its transformation interpolations. That is, **Ω** = 0.0

Finally, if the distance separating the object from the viewer is within the level of detail range, then the object dynamic ratio is computed as follows:

$$\mathbf{\Omega} = \frac{\boldsymbol{Current\ Distance - Minimum\ Distance}}{\boldsymbol{Maximum\ Distance - Minimum\ Distance}}$$

It ensures that the closer an object is to the viewer, the greater its dynamic ratio will be. On the other hand, its dynamic ratio will converge to 0 as it goes away from the viewer. With the per-frame dynamic ratio of each animated object calculated, we can finally compute each bone's current level of detail by scaling its current frame's static priority based on its dynamic ratio.

For each bone: i

Get current key frame: j

Get static ratio of key frame j: P_i^j

Scale static priority by dynamic **Ω**

Current LOD = P_i^j x **Ω**

As the final step, each bone's current level of detail is calculated with consideration of its current key frame, its static priority and the object dynamic ratio. This allows to manipulate the number of steps this bone's transformation will be interpolated.

Remember that, in order to decrease the computation time when updating the animated model, some of its bones' transformation interpolations are to be skipped according to the bones current static priority (based on current key frame) and the object dynamic ratio (based on object's distance from the viewer). An articulated model is updated based on the time elapsed from the last time it was updated. Therefore, this level of detail algorithm uses the time factor in order to decide if the current transformation interpolation of a certain bone should be skipped. The final priority of the bone represents the amount of time the update algorithm waits before actually skipping this bone's transformation interpolation. In other words, the current level of detail value of each bone will be used as the new time step for updating that bone's transformation matrix.

Example: Comparing 2 bones:

Current level of detail of B_1 is 0.048

Current level of detail of B_2 is 0.032

Notice that at run time, each bone is independent from all other bones, and it will skip its transformation interpolations using its own level of detail value. In the example above, the application is assumed to be running at 60 frames per second, therefore the duration of each frame is 0.016 seconds or 16 milliseconds. Since Bone B_1 has a higher priority, it skips an interpolation each 48 milliseconds, comparing bone B_2 with a relatively lower LOD value, which skips an interpolation each 32 milliseconds.

7 TEST AND BENCHMARKS

7.1 Performance gain

To verify the result of the animation LOD technique, the following test is conducted with 100 animated models, each made out of 153 bones and positioned in uniform grid. The level of detail distance range for this test is [200 – 1700]. That is, no level of detail will be applied to any model whose distance to the viewer is less than the range's lower bound (200 units). On the other hand, if that distance is greater than the range's upper bound (1700 units), then the model will skip all its transformation interpolations. Finally, if that distance is within the level of detail range, then a dynamic ratio will be computed and multiplied by each bone's current priority in order to generate that bone's current level of detail.

Note that when no LOD is applied, the frame rate is always fixed (At 12 frames per second in this test), because each of the 100 animated model is interpolating all its bones' transformations. However, when the application is set to take advantage of the animation level of detail optimization technique, its frame rate starts increasing when the distance separating the camera from the animated models becomes greater than 200.

An extreme gain in performance is noticed when the distance separating the camera from the animated objects becomes greater than 1400~1500. The one and only reason behind this fact is that at that distance, some of the objects are beyond the upper bound of the level of detail range and therefore all their transformation interpolations are skipped. The following figure illustrates the performance gain and comparison with the animation LOD approach.

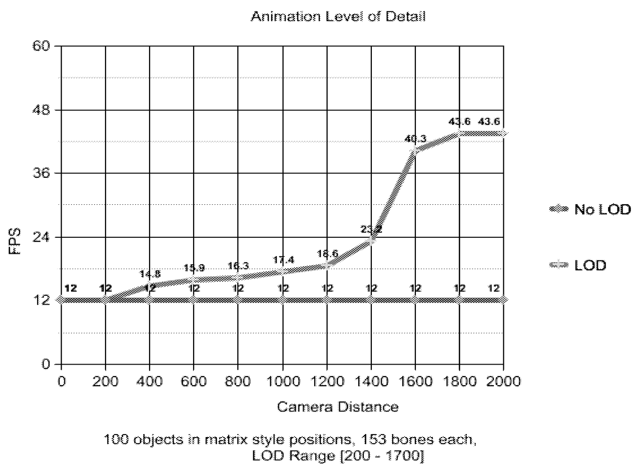


Figure 8: Frame rate increases when the distance separating the objects from the camera increases

In the next test, we use 100 of same animated models and place them at random locations in the scene. The same LOD range of [200 – 1700] is applied and the frame rates are recorded when the camera is zoomed in and out.

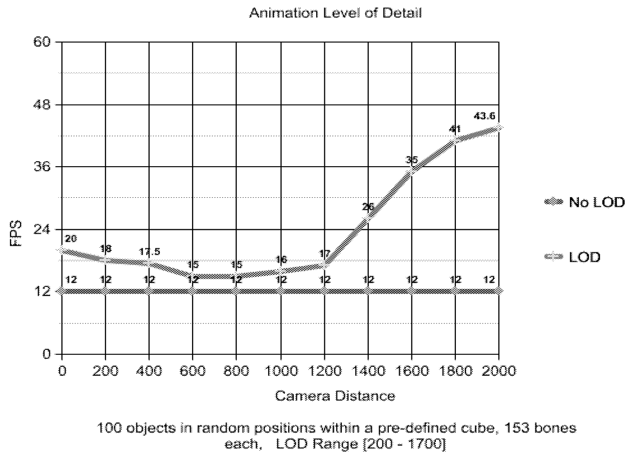


Figure 9: Objects are positioned as a cube

We can see that in this formation, the frame rate of the application is always above the normal frame rate where no level of detail is applied. This is due to the fact that some of the animated models are already far from the camera. Notice that the frame rate drops a little bit when the distance separating the camera from the center of the group increases. It is logical since at this point more objects are becoming visible. Animated models skip all their transformation interpolations when they are visible from the viewpoint. However, as the distance increases, we can see the frame rate picking up again, until it reaches its maximum value of 43.6 frames per second, which is identical to the previous test, because at this distance, all the animated models are skipping all their transformation interpolations.

7.2 Visual artifacts

Artifacts start appearing when a bone's level of detail is really low, which means lots of transformation interpolations will be skipped. The problem is mostly noticeable when a bone's animation doesn't get interpolated for a while, while its children bones are. This leads to a weird animated character that is noticeably different than what the animators had in mind.

In the next few screenshots, a debugging camera is used in order to better analyze the artifacts. The real camera distance, which is used by our LOD algorithm to determine the level of detail of each bone, is printed in the top left corner of the window. LOD is only applied to the character that is rendered on the right side

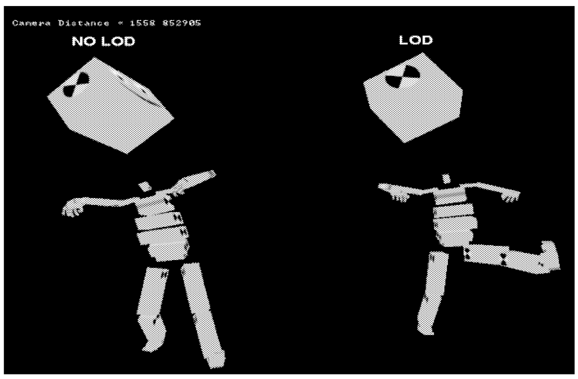


Figure 10: On left- lagging, on right:-correct

In Figure 10, the distance separating the camera from the objects is greater than 1550, while the LOD range is [200–1700]. It means that most LOD values are low and they're skipping many animation interpolations. We can see that the left leg of the character on the right is lagging (affected by our LOD algorithm), while the left leg of the character on the left shows the correct posture.

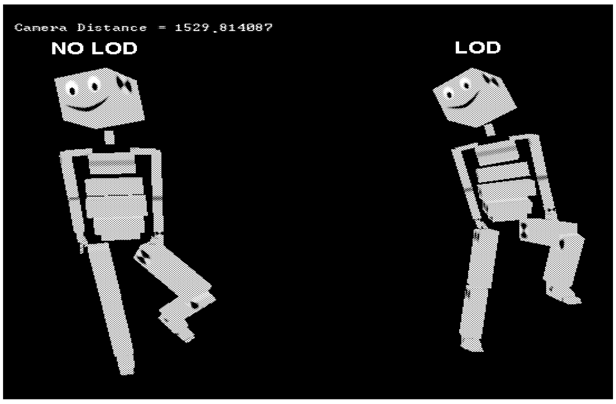


Figure 11: Both thighs' are lagging with different error margins

In Figure 11, the distance separating the camera from the objects is still great at 1530, and the same LOD range is used: [200 – 1700]. We can see that the left legs of both characters are lagging but with different error margins.

The previous artifacts are apparent mostly when the object affected by our LOD algorithm is being rendered without taking the distance separating it from the camera into consideration, and this is basically what was done in Figure 10 and 11. A debugging camera was used in order to render the objects up close, while the real camera was only used when computing the LOD values of the characters' bones. If the real camera is used for rendering, which is the case in any game and application, both these characters would end up occupying a very small screen surface, thus making these artifacts unnoticeable.

Figure 12: Characters rendered using the real camera

In Figure 12, the real camera was used to render both characters. Although the previous artifacts shown in Figure 10 and 11 are still there, the objects are barely visible up to the point where the artifacts are any noticiable.

8. FUTURE WORK

This level of detail approach for interpolated animations of articulated models presents a technique for reducing the amount of time needed to update animated models with minimal or no visual artifacts. The reduction decision is based on the current priority with six sub-components.

Currently these 6 sub priorities are treated equally, translating to the fact that they all contribute to the final priority value using the same weight. Alternatively, they can also be arithmetically generated by analyzing the model and collecting information on how much transformation parameters, depth and number of children bones contribute to the shape of the animated model. Base on this profiling, the level of detail algorithm can assign a different weight for each of the sub-priorities. The authors believe that it will enhance the adaptability of this LOD approach and therefore produce the better visual results.

REFERENCE

[1] Julie C.Xia, Jihad El-Sana, Amitabh Varshney, "*Adaptive Real-Time Level-of-Detail-based Rendering for Polygonal Models*" – IEEE Educational Activities Department, 1997

[2] Hoppe, Hugues, "*Smooth View-Dependent Level-of-Detail Control and its Application to Terrain Rendering*" - Microsoft Research, IEEE Computer Society Press, 1998

[3] Peter Lindstrom, et al, "*Real-Time, Continuous Level of Detail Rendering of Height Fields*" – ACM, 1996

[4] Markus Giegl and Michael Wimmer, "*Unpopping: Solving the Image-Space Blend Problem for Smooth Discrete LOD Transitions*" – Institute of Computer Graphics and Algorithms – Vienna University of Technology, 2006

[5] G. Schaufler, W. Sturzlinger, "*Generating Multiple Level of Detail from Polygonal Geometry Models*" – Springer-Verlag, 1995

[6] Hugues Hoppe, "*Progressive Meshes*" – ACM, 1996

[7] Mark Eastlick and Steve Maddock, *"Triangle-Mesh Simplification using Error Polyhedra"* - Journal of Mathematical Modeling and Algorithms, 2006

[8] Hiroaki Date, Satoshi Kanai, Takeshi Kishinami, "*Mesh Simplification and Adaptive LOD for Finite Element Mesh Generation*" - Ninth International Conference on Computer Aided Design and Computer Graphics, 2005

[9] Reinhard Klein, Gunther Liebich, W. Straßer, "*Mesh Reduction with Error Control"* - IEEE Computer Society Press, 1996

[10] Hung-Kuang Chen, et al., "*Generating high-quality discrete LOD meshes for 3D computer games in linear time*" - Springer Berlin / Heidelberg, 2006

Antoine Abi Chakra is a faculty member of the CS department. He earned his B.S. degree in RTIS and M.S. degree in CS both at DigiPen, where he now teaches algorithm analysis, game implementation techniques, fundamentals of game development and portable game development. Antoine also works with an R&D team at DigiPen on various projects and is responsible for designing, implementing and delivering solutions.

Xin Li is a full professor in the CS Department at DigiPen. He has a Ph.D. degree in Computer Science with the expertise in 3D computer graphics, real-time rendering techniques, animation and modeling algorithms. Prior to his teaching position at DigiPen, Xin had six years work experience at Lockheed Martin and three years at Nintendo Software Technology, writing flight and ground vehicle simulation and video game software.

ELEMENTS OF A CHINESE LANGUAGE GAME

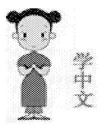
Anton Eliëns [1,2]

[1] Creative Technology, University of Twente
[2] multimedia @ VU University Amsterdam

KEYWORDS
chinese language, serious games

ABSTRACT

In this paper, we elaborate on work presented at the Dutch crossmedia PICNIC festival, in a special symposium entitled: the *China Language Bridge*. We will discuss a number of online resources, including games, for learning the chinese language, including chinese characters which are also used in japanese. After a brief digression on potential pitfalls in using online translation services, we will present some ideas, and indicate in what way these ideas might be realized using the XIMPEL interactive video platform, developed on top of the open source flex/as3 SDK. Such an approach would cover both the need to introduce appropriate context to stimulate language learning, as well as the unaviodable repetitions, which often proves to be one of the main obstacles for effective language learning.



INTRODUCTION

引言

At the PICNIC crossmedia festival[1] 2007, Chinese Radio Amsterdam[2], currently also known as CRTV[3], a local media organization for the chinese community in the Netherlands, organized the *China Language Bridge*[4], a symposium with a great variety of speakers, from a commercial as well as educational background. Topics covered by the symposium included: city and language guides on mobile phones, automatic translation of incoming and outgoing emails, and chinese language games.

There are multiple reasons to be involved in developing chinese language games, when looking at the chinese language in terms of global impact and the complexity, or for that matter intriguing nature of the language, especially from a western perspective, Ross & Ma (2006). Also, it should need no argument that games are an effective way to learn, Eliens & Ruttkay (2009), or at least support the learning of a language, taking into account that many elements are, in principle, available as online components, including grammar definitions and vocabularies. In addition, however, I must confess to my personal motivation to learn the chinese language. And, I must admit, it was much more difficult than I expected, giving me another reason not to give up. Learning to read, write and speak chinese, may the least of it considered to be an interesting challenge. And as such an excellent topic for game play!

structure The structure of this paper is as follows. We will first look at what constitutes a chinese character, and then briefly look at what online resources are available for (learning) chinese. As a digression we will present an online translation example, humorous in itself, but nevertheless a warning for potential pitfalls in this area. We discuss a simple language game, and introduce what in our opinion are core elements of a language game: play, learn and explore. Then we will propose a platform for the realization of (serious) language games, and present initial ideas for scenarios to be worked out in the future, after which we will draw (preliminary) conclusions.

By way of acknowledgement, all help in translation (帮助 in 翻译) is by google translation service[5].

WHAT IS IN A CHARACTER?

什么是一个中国人的个性？

For a western person, chinese characters are deeply intriguing, whether on the billboards of Shinjuku (Tokio) or the shops of Shanghai. Due to their shape and the implied meanings of the characters, chinese writing is significantly distinct from western writing, leaving the uninitiated spectator utterly puzzled. An interesting approach to teaching westerns the contruction rules of chinese characters is taken by the artist Xu Bing, who showed, and demonstrated in workshops, how to

[1] www.picnicnetwork.org
[2] www.chineseradio.nl
[3] crtv.nl
[4] www.picnicnetwork.org/page/5684/en
[5] www.google.com/translate_t?langpair=en|zh-CN

create words in western alfabet using elements and construction principles from chinese writing, as illustrated below, showing a possible logo for *chinese radio.*

www.xubing.com / 华语电台.nl

When learning chinese characters, above all, it seems important to understand the significance of the individual elements of which the character is constructed, and the rules of composition, that is the way meaningful combinations are made.

genealogical chart(s)[6]

1. pictograph(s) – e.g. 木 = *tree*
2. ideograph(s) – abstractions, e.g. 一 = *one*
3. logical aggregate(s) – e.g. 安 = *peace* (roof, woman)
4. phonetic complex – e.g. 忠 = *loyal* (center, heart)
5. associative transformation(s) – concept extension
6. borrowing(s) – unrelated, similar pronunciation

In addition to an explanation of the meaning of the elements and rules of composition, most textbooks dealing with chinese characters suggest the reader one or more metaphors to help memorizing the characters, which is, for example, easy to find in the character for *peace*. Cf. McNaughton (2005), Wang (1993).

RESOURCES FOR CHINESE

中文资源

There are ample online resources for chinese, including dictionaries[7], tools for reading and writing[8], the already mentioned google translation services[9], as well as overviews of learning resources[10]. Interestingly, many of these online resources can be accessed as a web service using a simple REST API, an explanation of which falls outside the scope of this paper.

With the rise of the mobile phones, gadgets such as the iTravl[11] are increasingly being replaced by applications for the mobile device. In particular, one of the technologies that may have a significant impact on mobile city and laguage guides is the *augmented reality* browser for iPhone and Android from Layar[12].

[6]zhongwen.com
[7]www.mdbg.net/chindict
[8]www.mandarintools.com/dimsum.html
[9]www.google.com/translate
[10]www.chinapage.com/learnchinese.html
[11]www.ectaco.com/products/ECTACO-TL-2EuAs10
[12]layar.com

Fig. (1) iTravl: handheld translator

As a note, a reference worthwhile for the ardent LaTeX/ TeX users is provided in the instruction pages for the CJK[13] package.

an exercise in online translation To illustrate potential pitfalls of automated translation, look at the fragment(s) below.

daft punk – technologic

Buy it, use it, break it, fix it.
Trash it, change it, melt – upgrade it.
Change it, point it, zoom it, press it.
Snap it, work it, quick – erase it.
Write it, out it, paste it, save it.
Load it, check it, quick – rewrite it.
Plug it, play it, burn it, rip it.
Drag and drop it, zip – unzip it.

Translating this into chinese, omitting the last 4 lines, gives us:

chinese version

买它,使用它,突破它,把它修复.
垃圾它,改变它,融化–升级.
改变它,它一点护着它,它的新闻.
折断它,它的工作,快速–抹掉.

Now, translating it back to english brings distortions to the surface which, humorous in this context, might be devastating in a more serious application.

after translation

Buy it, use it, break it, fix it.
Refuse it, change it, the melting – escalation.
To change it, it is 1:00 and protecting it. it is the news.
Breaking it, it's work, fast – erase.

Admittedly, this is not a representative example, and, to be honest, I am impressed by the overall quality of translations as provided by for example the google translation service. However, such errors may also be used in a language learning game, as a (language) mini game about ambiguity of meaning, in particular in chinese where for example mispronuncations, which are quit common due to the unfamiliar tonal character of the language, may lead to widely divergent meanings.

[13]www.math.nus.edu.sg/aslaksen/cs/cjk.html

ONLINE LANGUAGE GAMES

在线游戏语言

A simple selection game, for memorizing kanji (chinese characters used in japanese) can be found in the *online kanji game*[14], which allows to switch between all combinations of english, kanji and kana (the native japanese characters). A prototypical language learning game is knuckles in china land[15], also for japanese, which takes the player through various rooms and requires the player to answer (in a very simple way) to particular situations and challenges.
To support a television program for learning chinese, the dutch educational broadcasting society provided a website with additional online games[16], mainly consisting of images of situations or room, where the player must guess, after some training sessions, the meaning of objects, by selecting the appropriate chinese word.

Fig. (2) I go to Shanghai for ...: 我去上海出差

In a similar way, many chinese learning tools provide scenarios, in which the user/player takes a particular role in a prototypical situation, as indicated in figure 2.

PLAY, LEARN, EXPLORE

游戏内容–玩,学习,探索

To communicate in a foreign language, in particular for young children, nor grammar nor an extensive vocabulary are important. Most important is that a limited set of words can be used in an appropriate context.


Fig. (3) Learn by Play

[14]www.msu.edu/∼lakejess/kanjigame.html
[15]www.kicl.info
[16]www.teleac.nl/chinees

Figure 3 illustrates a simple game to learn the words for the various colors, simply by stepping on the right spot on the (color) mat. Either in real life or by means of video, such situations may be helpful to acquire intuitive mastery of such canonical vocabulary as colors, or for example, greetings, or even table manners.
Such games provide a suitable context, as well as a task that is entertaining in itself, thus motivating the player to repeat, that is to jump to the right spot whenever a new color is called for.
In general the (game) mechanics underlying such interactions may be summarized as:

game(s)

- play – turn-based, score(s)
- learn – simulation model, target(s)
- explore – interactive video, mini games

Such (game) mechanics, as illustrated in figure 4, allow for dealing with the various elements of learning a language, that is: providing a natural context with proper challenges, well-defined tasks and repetition, repetition, repetition.

SERIOUS GAMES PLATFORM

一个平台,为严肃游戏

In Eliens et al. (2008) we have described a platform, originally developed in the context of a climate game, Eliens et al. (2007), that allows for a seamless integration of the various elements of a (chinese) language game, as indicated above. See figure 5.

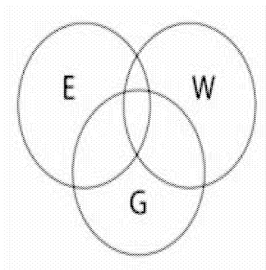


Fig. (4) Game Play, Simulation, Exploration

Technically, the characteristics of XIMPEL may be summarized as:

XIMPEL[17]

- ajax – dynamic update(s)
- flex/as3 SDK – XML game description template(s)
- (flash) video – interactive video, mini game(s)

However, more importantly is that XIMPEL has proved itself to be a viable platform for media productions that fall anywhere between interactive storytelling and gameplay, and also allows for tracking choices of the user/player to provide a meaningful interpretation of the score, Eliens & Ruttkay (2008).

[17]ximpel.net

STORY & GAME SCENARIOS

故事和游戏场景

Having taken an active interest in learning chinese, and chinese culture, I would also like to suggest particular topics around which chinese language games could be developed.
One (obvious) topic is kung fu, written as 工夫 , the literal translation of which is *hardworking men*, not excluding women by the way. The literature abounds with stories and myths of practitioners of martial arts with superhuman powers, in search for missions to prove their mastery and solve human problems. CF. Liang & Wu (1964).
Another line of stories may be derived from *Wei Cheng* (围城), the famous novel of the writer and scholar Qian Zhongshu (钱钟书), which relates about the home coming to Shanghai of the main character Fang Hongjian (方鸿渐), who will later leave Shanghai for an eventually failing career at a university in mid-China (三间大学). The opening sentence of this novel (in the learners edition) reads:

围城

红海早过了，船在印度洋面上开驶着，但是太阳依然不饶人地迟落早起，侵 占去大部分的夜。

Apart from being an excellently written novel, the life of the main character bears some resemblance to my own life. Moreover, the background of the story, Shanghai in a period of global turmor provides an excellent stage for in some sense the anti-hero Fang Hongjian, thus allowing for a range of quasi-dramatic interactions and plots. Even in modern times, this may lead to compelling cinematic games, with questions such as *what are you going to do in Shanghai* (figure 2).

CONCLUSIONS

结论

In this paper we have investigated some of the possible elements of a chinese language game, both from a personal interest and technical interest, and, most importantly, from a more general interest in game development as well. Written from an entirely western perspective of chinese language learning, it may nevertheless give the inspiration and some insights of how to develop a chinese language game, making use of online resources and web services as they become available.

ACKNOWLEDGEMENT(S) Thanks to Hong Tong Wu for involving me in the Chinese Radio Amsterdam activities, and inviting me for the PICNIC 2007 event. And to my wife Yiwen Wang, not for teaching me chinese, but above all for making not only the chinese culture but everything else as well so vibrantly alive!

REFERENCES

述

Eliens A., van de Watering M., Huurdeman H., Bhikharie S.V., Lemmers H., Vellinga P. (2007), Clima Futura @ VU – communicating (unconvenient) science, In Proc. GAME-ON 07, Bologna, Italy

Eliens A., Huurdeman H., van de Watering M., Bhikharie S.V. (2008), XIMPEL Interactive Video – between narrative(s) and game play, In Proc. GAME-ON 2008, Valencia, Spain

Eliens A. and Ruttkay Z. (2008), Record, Replay & Reflect – a framework for understanding (serious) game play, In Proc. Euromedia 2009, Brugge, Belgium

Eliens A. and Ruttkay Z. (2009), Math game(s) – an alternative (approach) to teaching math?, In Proc. GAME-ON 2009, Dusseldorf, Germany

Linag Shou-Yu and Wu Wen-Ching (1964), *Kung Fu Elements – Wushu Training and Martial Arts Application manual*, The Way of the Dragon Publishing

McNaughton W. (2005), *Reading & Writing Chinese Characters – Simplified Character Edition*, Tuttle Language Library

Qian Zhongshu (1947), *Wei Cheng (The Besieged City)*, Sinolingua 1994 (student edition)

Ross C. and Ma Jing-heng Sheng (2006), *Modern Mandarin Chinese Grammar – A Practical Guide*, Routledge

Wang Hongyuan (1993), *The origins of chinese characters*, Sinolingua

Anton Eliens is professor creative technology / new media at the University of Twente and coordinator of multimedia @ VU University Amsterdam. He has experience in web-based media, interactive video, and the development of serious games.

COMPLEX SYSTEMS SIMULATION

ANALYSIS OF COMPLEX SYSTEMS MODELLING

Renate Sitte
Griffith University
SEET/ICT, Gold Coast Campus
Griffith University, QLD 4222, Australia
E-mail: r.sitte@griffith.edu.au

KEYWORDS
Complex systems, modelling, analysis, simulation

ABSTRACT

With increased applications of complex systems the need to develop an overarching theory, a methodology and good practice becomes increasingly urgent. This paper provides guidelines for advanced techniques that aid in the systematic analysis of complex systems in a multidisciplinary environment. The results of the analysis are suitable stepping stones for the design and development of simulation and prediction models. First we look at fundamental complex system analysis guidelines, and then then we move to a selection of specific practices that are particularly helpful in the analysis complex systems. Examples from simulations in various industries are presented.

INTRODUCTION

Designing a new mathematical model to simulate a complex system can be as challenging as finding a path in a labyrinth being blindfolded. This is because we have to model and simulate the very processes which we are trying to understand. This means modelling the unknown *while* getting to know it. With steadily increasing computing power and the development of computerized tools with better simulation capabilities, the study of complex systems modelling has rapidly increased, but the progress of theoretical scaffolding has lagged behind. The development of an overarching theory of complexity is urgently needed. To achieve this, the first step is to provide guidelines of good practice, to strengthen our understanding from which a theoretical framework can emerge. In our earlier work (Sitte 2009) we have provided an insight of how to approach complex systems in general and identified the need of a unifying definition of complexity that can lead to its quantitative measure. This paper builds on our earlier work providing guidelines for advanced techniques that aid in the systematic analysis of complex systems in a multidisciplinary environment. The results of the analysis are suitable stepping stones for the design and development of simulation and prediction models. We look at fundamental complex system analysis guidelines, and move then to specific practices that are particularly helpful in the analysis of complex systems.

There are several ways of approaching the study of a complex system. One of them are the holonic systems (Koestler 1967) which triggered the rise of many other developments, among them the multi-agent systems. In this school of thought, a holon can be both, an autonomous system, and a functional part of another system, where a holon has the potential to be in a state of order or chaos, that is, its behaviour is predictable, either independent or depending on its initial conditions. Another variation of the holonic approach is the spiral system (Beck and Cowan 1996), which includes evolution. However, all these systems in their current forms are of a rather philosophical nature. They do help us to get our ideas in order, but their help in setting up a model that allows us to analyze, predict or engineer a complex system is somewhat limited to specific cases. With systems theory no longer sufficient for larger and more complex systems (Bar-Yam 2003) we have to find new ways for dealing with complex systems for their understanding and design, and appropriate simulation environments to predict their behaviour.

THE ANALYSIS OF COMPLEX SYSTEMS

First of all, a model is not, and will never be a realization. The model will always be, at best a reasonable approximation of what it intends to simulate, regardless whether it is for observation and study, or for engineering and design purposes. Some models require networks of ambitious sets of equations for comprehensive simulations, while others are limited to perhaps a few differential equations that model just a single aspect of a system. More equations, more computations do not necessarily make better simulations. As always, more computations mean more chance for error.

In designing simulation models for complex systems, we must have a clear idea of what we are after because depending on the purpose of the study, this will affect the depth and ramifications of the approach. There is a big difference in trying to establish a model for something that we want to do, and what we *can* do, as for example the design of a flexible manufacturing factory as opposed to designing climatic changes. The former we can change and construct, the latter we can only observe, understand, and predict but its design is yet limited by technology state of the art. The models for simulation suitable are and must be different depending on the purpose. They are not only a matter of the intention but also of the ability to accomplish that intention with current means.

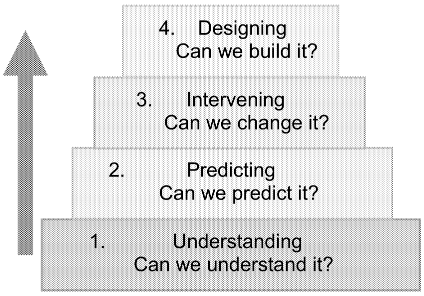


Figure 1 Basic questions for analyzing the purpose of modelling a complex system

We have to distinguish whether we are able to intervene and change the system. The fundamental questions to ask are:

1. Do we want to learn or understand the system? This would be the case of social and economic dynamics.
2. Do we want to predict the system? For example meteorological and political systems.
3. Do we want to intervene and affect the behaviour of a system? For example strategies in national or international socio-economic environments.
4. Do we wish to engineer and design a system. For example in genetic manipulations for vaccines or power for a community supplied from different sources (hydro, wind, solar, etc).

At first these questions might appear as trivial, even equivalent. They are not. Figure 1 shows the required precedence of the purpose of a complex system model. The order of these questions is escalating. This means, the answer to a question relies on being able to answer affirmatively the preceding one.

The next set of questions that we have to answer is whether we will *be able* to do what we expect, because this cannot be generalized for any situation. For example, let us take turbulences which are known to be difficult to model. We cannot act on atmospheric turbulences, but with clever design, we can create turbulences in microsystems (MEMS) to mix microscopic quantities of fluids for applications in medicine. We might be able to understand the dynamics of stock markets or whole economies (question 1), but we might not be able to stop them from crashing like dominoes as they did in 2009 (question 3 or 4). However, what we are unable to do nowadays could be possible in the future with the rise of different technologies or knowledge. This is part of the nature of complex systems, and it goes back to the notion that complexity is not necessarily always an inherent attribute of the system itself, but it is also part in our inability to understand or conceptualize them. This can be lack of knowledge, maturity of knowledge (to "see" something). For example, we can predict a volcanic eruption, but with current knowledge and skills we are not able to intervene. We might want to design a vaccine but unless we understand *what* constitutes the lever effect in making that vaccine the means for whatever the immunization case is, we will most likely get an unexpected outcome, which could be as little as minute differences in initial conditions.

The next step of the analysis is to learn more about the system in a focused way. Again this might sound déjà vu, but it is not for complex systems. Some of the initial approaches were presented in (Sitte 2009). We support the stance that not all in a complex system is complex. We have to find out *what* is it that makes the system complex. In many cases, but not all, this follows an 80/20 rule (Pareto principle). A corollary of it is to find those 20% of the entities that make the complex system to be complex.

To distinguish in what the complexity consists, we classify into the fundamental groups of complexity, which were presented in our earlier work, but for the sake of completeness and help in understanding what it is about, we repeat them here below and show its interrelations in Figure 2. We analyze different aspects of the complex system.

To do this we look at the functionality, structure, topology, and architecture. This scheme is not exhaustive and could be extended, however it shows the fundamental aspects to make the process understandable. The reader should also consult the original publication.

(i) Structural (dimensionality, networks, hierarchy with its subclasses breadth and depth)
(ii) Functional (single or multifunctional components)
(iii) Topological (connectivity, relations with its subclasses number and direction) , and
(iv) Architectural (series, parallel, cascade with subclass recursion).

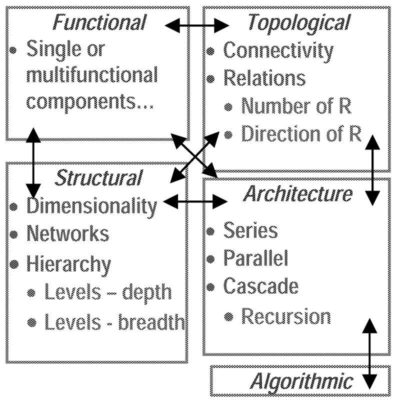


Figure 2 Fundamental types of complexity and their affinities (Sitte 2009).

A word of caution: the concept of complexity theory as used in computing (the "big O"), refers to algorithmic complexity. Its applicable to other complex systems is very limited.

DEFINING BOUNDARIES AND FINDING VARIABLES

Once we have sorted out what the purpose of our study is, by defining the intention of observing or intervening into the complex system under investigation, we will be able proceed to the two familiar steps that we have used so many times before: (a) defining the boundaries of our complex system and (b) defining the variables.

Defining the boundaries is made easier by using the four questions presented above. This chain of thought is important because, as in any system modelling, complex or not, it influences which variables are necessary to describe the system. One problem in modelling complex systems is that sometimes we do not know which variables are relevant and which ones are not. Indeed, sometimes we do not even know which magnitudes can become a useful variable for setting up a model. Intuition is not a reliable advisor, in particular when the system to study can have chaotic behaviour, perhaps clouded by other information.

Let us take the example of a transistor. For the past 50 years transistors have been the core of our fast evolving computing industry. The fabrication of a transistor requires fine tuning of over 200 processing parameters, whose manufacturing fluctuations affect the dimensions and functionality of a transistor. For many decades, it was believed that most of its 200 parameters were influential. Transistor design was based on experimentation to derive a specific manufacturing recipe. It was not until 15 years ago that a comprehensive sensitivity analysis revealed the concept of a multidimensional hull of processing-parameter values within which a functional transistor could be warranted, and be utilized in design trade-offs resulting in different performances. Surprisingly, only a few (less than 20) parameters, of the 200 parameters had a strong influence to be used in design engineering (Sitte 1994). The example illustrates how important it is to rely on sound analysis instead of human belief.

VARIABLES, MAGNITUDES AND SCALES

The magnitudes that will be part of our model must be observable and measurable. Sometimes the observations and measurements have to be done in an indirect way, sometimes they can only be estimated. This can be difficulted by changing values in time as in dynamic systems. However, in complex systems the dominant variables (not just their values) that are critical to performance can change to other variables in different conditions. To focus on the effect of changes on the behaviour or functionality is usually a good start. There are a number of traditional ways to approach this, such as a combinatorial sensitivity analysis. While the "combinatory curse" has been overcome by computational power, there are still cases of intractable computations such as manybody interaction e.g. interplanetary. In the development of mathematical models for complex systems, the main problem is that a sensitivity analysis is not possible without a mathematical model of at least part of our complex system. In pioneering endeavours to develop a model where the p is scarcely known or difficult to explore, this becomes hen-egg dilemma.

An increasing popular technique for identifying critical variables for modelling and design is the use of *rough sets*. Introduced in the 1930's this is a rather recently revived paradigm with its own theory; it is suitable for processes or data affected by uncertainty. Rough sets are used in classification, decision tables and data mining and granular programming. They are known to be able to extract laws from data, and are often applied in conjunction with fuzzy sets. The essence of rough sets is the grouping into classes and removal of redundant information (attributes) leaving the *reduct*. This is a subset that is the minimal set containing the information to characterize it as a class. A special form of reducts is able to recognize patterns in chaotic (unstable) data (Bazan et al. 2000).

Another important aspect to consider in the analysis of complex systems is structural stability, that is, the robustness that the performance of a system to internally caused disturbances. This is also analyzed using in Rough Sets paradigms (Andronov and Pontryagin 1937). Structural stability is relevant in the distinction between the choice of a functional design and more robust equally functional design that is less prone to instabilities.

Here is an example in the design of micro-electro-mechanical systems (MEMS) that illustrates the possible problems with apparent functional stability. In MEMS design our intuition from macroworld mechanics cannot be applied. This is because in the microworld different physical forces become dominant. In the macroworld, inertia and gravity are dominant, but in the microworld these forces are literally overpowered by adhesion, friction and capillarity, which are dominant in MEMS. If we were to design a pump by downscaling a macroscopic pump design, it would fail to pump properly due to the different physical effects that become manifest and dominant in microscopic scales. The functional stability that exists in macroscopic pumps fails, and is displaced in the microscopic pump by internally caused turbulences caused by friction (electric double layer) that appears at high velocity in fluids (slip length). In the microscopic design, turbulences arise in smaller channels (approx. 50 μm) while in larger channels its effect is nonexistent. The exact critical dimensions depend on other geometries and the fluid to be pumped but its criticality is in the order of 1 μm or less.

This example illustrates the importance of examining the internal stability of apparently equal configurations of systems, and their vulnerability to a minute change of the dimensions of the device under construction. In a similar situation and with the help of rough set theory, equivalence classes could be found and identify those with better internal stability for a robust design. But this requires real data from existing devices.

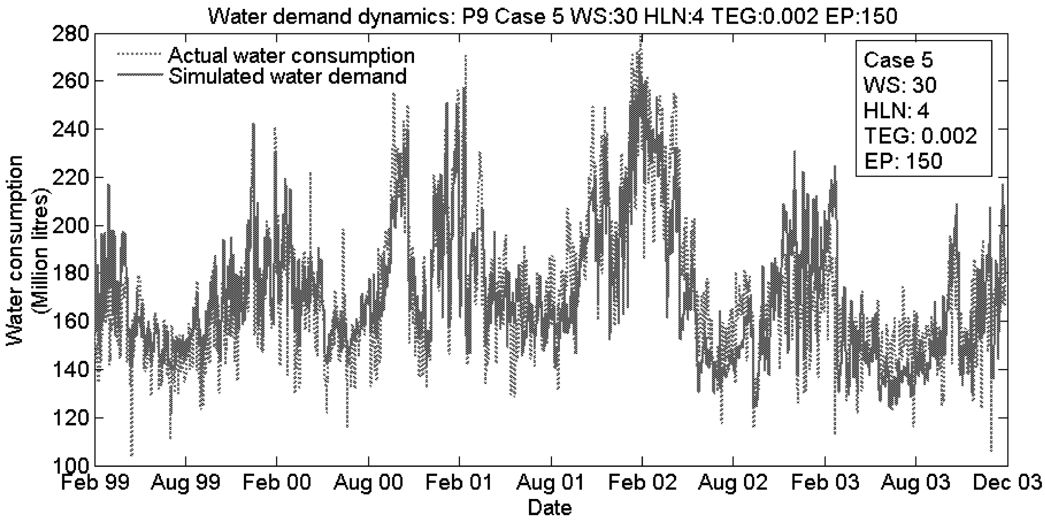


Figure 3 Validation of the NN performance compared with earlier year's actual data. The NN was trained with data sets of different years than shown. The NN has picked up the regular cycles of increased water consumption in the summer months Jan-Feb. It also shows that the NN model slightly overestimates the consumption. Source: (Udono 2005)

THE RECOGNITION OF STRUCTURES AND PATTERNS

One might easily believe that a system with many components that are heavily inter-related is more complex than systems with fewer components or less interrelations. However the simplicity of the equations that generate Mandelbrot sets are a good counterexample to illustrate how easily our intuition can fail.

An important part of the analysis for deriving a model for a complex system is to find its patterns, be it behavioural, modal, functional or other. If there are any, we have to find out whether these patterns are recurrent, cyclic or not. However, it can be difficult to find patterns in a first instance, although they might seem to exist, e.g. in a chaotic system. Again, intuition may or may not be a reliable aid. How can we find patterns when we have no mathematical model to replicate or study such patterns? Artificial Neural Networks (ANN) are a good tool to find out behaviour of a system that has input and output. There is ample history in the literature where ANN have been applied successfully to solve a range of problems, and other times they seem useless and overly abused.

The problem with ANN is that they do not maintain a mapping with the internal mechanisms that are involved in the process that they mimic; their internal processing is different. ANN are capable of learning and detecting behaviour in sequences of data, as a relation between input (data) and output (behaviour), but they do not provide an explanation of *which* variables or mechanisms of interactions produce the output like physical models do. While this appears as a disadvantage, it is an advantage for finding patterns in the behaviour, given the appropriate input.

DETECTING PATTERNS AND CYCLES WITH ANN

The types of artificial neural networks that have been most widely applied are the feed-forward networks. They are able to classify *static* patterns, or to approximate time independent functions. A more advanced group are the ANNs that are capable of representing *dynamic* systems. They are networks with internal feedback, the so called recurrent neural networks. However the understanding of recurrent neural networks lags far behind of our understanding of the feed-forward networks. By structuring the input to feed-forward NNs (static) these can be applied to time dependent patterns (dynamic); they are called Time Delay Neural Networks (TDNN). TDNNs are a class of NN that is suitable to study time series. Their input patterns are sequences of data, whose size represent a window into the time dependent data stream. The window is made to slide over the time range in discrete steps. At each position the data in the window form a "static" input pattern to the NN that characterizes the time changing data around a point in time. This corresponds to higher derivatives at a data point. An example would be a window with rising, falling, peaking, etc. data values. TDNN (feed forward) are popular NN configurations that are relatively easy to use, with ample software support available.

For a successful performance, the configuration of the NN has to be suitable to the data, and this requires several trial runs changing the number of internal nodes, the number of layers, and the window length until an acceptable network is found. Parameters for success are the mean square error. The idea is to train a network with progressive subsets of overlapping data (windows) and use a network configuration with as few as possible nodes. More nodes do not necessarily produce a better performance; they rather tend to overfitting of data. The range of applications for cycle detection using ANN is vast, and it is not the purpose to repeat them here or to provide a review of the use of ANN in complex systems. The purpose of the following examples is to illustrate the scenario of complexity of a case and what was possible to achieve with the ANN.

Application example 1: Pattern and cycle detection.

This example has been chosen for its illustrative potential of a partial model in a complex system used for a feasibility study. The example demonstrates how, in given situations, the choice of specific variables that are apparently critical can be left out, and successfully be replaced with just their behavioural effect in the form of a function determined by the ANN.

We have used ANN for a pattern and cycle detection as part of a project in modelling seawater desalination using clean waste-incineration energy (Udono and Sitte 2008). We used dynamic system modelling for this study to implement an overall comprehensive model. This model included the main cycles of water and energy and their modelling supply and demand for a fifty year span. The model included weather conditions which in Australia are affected by drought, flooding, and repeated El Niño phenomena; the model also involved population growth with an increasing but seasonally fluctuating water demand and unreliable supply by rain, and water restrictions.

Our specific case also required setting up a model for the supplying dam water dynamics that would have been far too complex and lengthy to implement using conventional physically based and statistical methods. This was compounded with problems of un-reliable data collection, and missing data which would have distorted the results.

The example presented here is about replicating non-linearity of dam level dynamics without going into laborious detail of conventional statistical models. ANN have been used for many years in studies involving the dynamics of water in hydraulic cycles and other courses, but never in the context as we used it. The aim was to train the ANN to compute a non-linearity of dam level changes. For example, the dried catchment due to long drought condition with less runoff into the dam, in contrast to the saturated catchment due to continuous rainfall causing higher rise of dam level. We used a feed-forward network with three layer architecture and an input combination of rainfall and water consumption. After training (and testing with past data) the mathematical model retrieved from the trained ANN was incorporated into the comprehensive (dynamic system) model. In this way we were able to use a working model but leaving the details of the true phenomena as a black box.

The result of the validation of the NN's performance is shown in Figure 3. The cyclic behaviour is consistent with the seasonal swings and the erratic occurrences of El Niño phenomena was captured by the NN. For the validation we used a set of data that was excluded from the training data set.

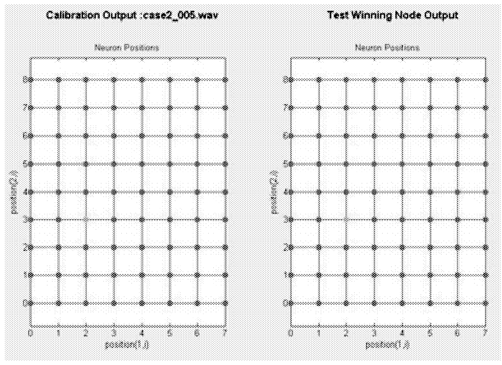


Figure 4 Example of a grid with 63 nodes to allocate 60 sounds. The node (3,2) is allocated to a specific sound (left pane), and successfully identified as winning node in position (3,2) (right pane).

DETECTING CLASS GROUPINGS WITH SOM

Self organizing maps (SOM) are a special class of NN. What distinguishes the SOMs from other NNs is that their training is unsupervised. This means that they produce their own mapping, as opposed to prescribed allocations in other NNs by data input/output pairs. The distinguishing characteristic of a SOM is that it preserves the topology of the input space. That is, inputs that are close (nearby) in the input space will result in activations of nodes that are also close on the grid of the SOM. This has lead to a range of applications, and particularly they are used for classification of objects. They are able to detect similarities in large data sets, and group these objects according to their similarities. Given a set of data they can detect clustering, or for example, the number of classes that the data can be grouped into. This in turn can be part of a simplification process, or aid in recognizing an internal structure.

Application example 2: Classification using Self Organizing Maps - SOM

This example is about environmental sound identification. Speech recognition, speaker recognition and recognition of music have been the focus of vast research. However, environmental sound recognition has been largely neglected until rather recently, despite its large potential for applications. Possible applications range from surveillance applications to detect intruders, to warning systems for the deaf, or remote carers of elderlies living alone. What makes environmental sound recognition so difficult is that (in theory) it comprises the totality of the sound spectrum, both in amplitude and frequency, with the potential of an infinite combination of sounds, while human speech, speaker recognition are limited to a rather small subset of what human vocal cords and specific instruments can produce. Techniques that have been traditionally considered as the best performers for speech/speaker or musical instrument recognition are either not suitable, or fail for environmental sound identification. This is due to the absence of a phonetic structure such as subwords or phonemes as is naturally found in human speech and the distinctive mark of "when does a sound start".

For many years environmental sound identification has not surpassed a 60 % success rate from a of sample space pool of sounds. We have been able to achieve a 90% success rate using SOM using unknown sounds in a pool of 240 different trained sounds (Sitte & Willets 2007). This is vastly more than the sound pools and experiments in any of the earlier reported papers. During the training the SOM sets up its own classification or grouping of sounds. After training, the SOM needs to be calibrated. This is a reverse engineering process, to identify which sound has been mapped to which node or nodes. It is necessary for validation and can be automated.

We used a two step identification process, but this is not always necessary. The first step identifies a cluster of sounds which also contains the sound to be identified. The cluster is then analyzed again, which regroups and isolates the sound to be identified. The result is shown as a green dot on the winning node in Figure 4. We do not know which other sounds were in each cluster, but it is possible to find out. It would have been an interesting but fiddly exercise of reverse engineering in a pool of 240 different sounds.

A more sophisticated variant of SOM is capable of modelling and classifying time dependent phenomena. The time-delay NN technique explained above can also be extended to SOM to let time dependent patterns "self-organise" according to their inherent similarities.

THE IMPORTANCE OF MEASUREMENT

Earlier we have stated that variables must be observable and measurable. While this is a well known factor in all scientific studies, it is important to be aware that the process of data collection, be it counting success or be it measurement of data, the reliability of the data has to be asserted and documented. Data can easily be distorted by incorrect measurement, by error from indirect measurement, and by the inherent error of the instruments that are used for the data collections and the limitations. For example, in the previous example of sound identification, the identification of sound is limited by the quality of the microphone that picks up the sound. When we are dealing with identifying the critical data required to describe a system, the inclusion or exclusion of a variable, albeit under uncertainty, can make the difference between complexity and simplicity at the time of modelling the complex system, between good model prediction and a bad one. This can be caused by ease or difficulty of mathematical manipulations at the time of deriving a mathematical model, or the similarity with known models, but it can also be caused by algebraic operations in the formulation and compounding error.

CONCLUSION

Setting up new mathematical models for complex systems that are suitable for simulations can be compared with modelling the unknown. In this paper we have presented the fundamental approach for the analysis that aids in complex system modelling for simulations. We have presented this in the form of steps and questions that help in choosing the quantities that are relevant in setting up modelling. We also have suggested advanced techniques that are suitable for complex systems and presented a variety of examples. This work is continuation of our earlier work. Future work is aimed at extending the modelling process, and with it, to enable a path towards an overarching complex system theory.

REFERENCES

Andronov, A. A.; Pontryagin, L. S. (1937), "Systèmes grossières", *Dokl. Acad. Sci. URSS (N.S.)* 14: 247–251

Bar-Yam Y. 2003. "When Systems Engineering Fails - Toward Complex Systems Engineering", *IEEE International Conference on Systems, Man and Cybernetics*, Volume 2, 5-8 Oct. pp 2021 - 2028

Bazan J.G., Nguyen H.S., Nguyen Ss.H., Synak P., Wroblewski J., "Rough Set Algorithms in Classification problem" in *Rough Set in Rough Set Methods and Applications 2000*, Physica Verlag (Springer Verlag), Eds. Lech Polkowski, Shusaku Tsumoto, Tsau Y. Lin, Heidelberg New York, pp 49-65, ISBN 3-7908-1328-1

Beck D., Cowan Ch., "Spiral Dynamics: Mastering Values, Leadership, and Change" 1996, Blackwell Publishing (Business), ISBN 1-55786-940-5

Koestler A., 1967. The Ghost in the Machine. London: Hutchinson. 1990 reprint edition, Penguin Group

Sitte R., "About the Predictability and Complexity of Complex Systems" in *From System Complexity to Emergent Properties*, 2009 M.A. Aziz-Alaoui & Cyrille Bertelle (eds), Springer series "Understanding Complex Systems", Part I, pp 23-48, ISBN 978-3-642-02198-5

Sitte R., Dimitrijev S., Harrison H.B., "Device parameter changes caused by manufacturing fluctuations of deep Sub-micron MOSFETs", ***IEEE** Transactions on Electron Devices, Special Issue on Fluctuation Phenomena in Devices,* Vol 41, No. 11, November 1994, pp 2210 - 2215

Sitte R., Willets L., "Non Speech Environmental Sound Identification for Surveillance Using Self Organizing Maps", *IASTED 4th International Signal Processing, Pattern Recognition, and Applications ISPRA*, Innsbruck (Austria) February 2007, pp 281- 286 ISBN 978-0-88986-646-1

Udono K., Sitte R., "Modelling Seawater Desalination Powered by Waste Incineration using a Dynamic Systems Approach", *Desalination* (Elsevier) 229 (2008) 302–317

Udono K. "Modelling Seawater Desalination with Waste Incineration Energy using Dynamic Systems Approach" Doctoral Thesis 2005 , Griffith University, Australia

AUTHOR BIOGRAPHY

RENATE SITTE was born in Vienna, Austria, and lived for longer periods in Venezuela, Sweden and Australia. She obtained her PhD (1995) and MPhil (1993) from Griffith University, Australia, in Microelectronic Engineering. She received her Systems Engineering degree (Ingeniero de Sistemas) in 1985 from the Universidad de Los Andes in Venezuela. She began her academic career as a faculty staff in 1986 in the early days of the Faculty of Engineering and Information Technology, Griffith University, Australia. She is now an adjunct Associate Professor in SEET/ICT at Griffith University, Australia. Her field of research is in modelling and simulation, with over 80 publications. She has supervised many postgraduate students, several in a major project in Virtual Reality Prototyping Micro-electro-mechanical systems (MEMS), which has provided and endless source for research topics.

OPTICAL FLOW MODELING: APPLICATION TO NAVIGATION OF A FIXED WING MAV IN A 3-WAY JUNCTION

Mahdi Rezaei, Fariborz Saghafi
Department of Aerospace Engineering, Sharif University of Technology
Azadi street, Tehran, Iran
Email: mrezaei@ae.sharif.edu, saghafi@sharif.edu

KEYWORDS
Optical Flow, MAV, Navigation, Collision Avoidance

ABSTRACT

Future micro aerial vehicles (MAVs) will have missions occur within and around buildings. This paper describes a vision-based navigation of a very light fixed wing aircraft in flight between obstacles. A geometrical method has been proposed to model the optical flow measurement process using the camera perspective model, the flight equations and the scene geometry. The optical flow information are used to detect the obstacles and perform turn maneuvers based on the balance strategy. The flight of an aircraft in a corridor which encounters a 3-way junction was simulated. The effect of the ways width ratio and the cameras' field of view angles on the aircraft navigation behavior at the junction are presented. The six DOF equations and complete aerodynamic data were used in aircraft flight simulations.

INTRODUCTION

MAVs are very constrained in weight and size and most of the conventional navigation sensors like laser, sonar or radar range finders which work effectively for ground-based robots are too heavy and large and energy consuming for MAVs. Optical flow techniques are one of the vision-based methods that mimic the flight control of flying insects (i.e. honeybees, flies, etc.) to create a feasible control system for MAVs. In these techniques there is no need to construct 3D model of the environment from 2D images. They use only qualitative information of images rather than their detail information, so the large amount of computations are avoided.
The utilization of optical flow in robots navigation has been interested since early 90s. Most of the efforts have been carried out on ground-based robots with missions inside buildings (Victor and Sandini 1995 ; Argyros and Bergholm 1999 ; Argyros et al. 2004 ; Camus et al. 1999). In the recent decade some efforts have been started about aerial robots and MAVs. Green et al. carried out some experiments on a very light aerial prototype using an optical flow sensor designed by Barrows (Barrows 1999). They investigated some control and navigational tasks like obstacle detection, autonomous landing, and altitude control by optical flow algorithms (Green et al. 2003, Green et al. 2004, Oh et al. 2004). Hrabar and Sukhatme used optical flow extracted from the images of cameras mounted on a RPH to navigate it through urban canyons (Hrabar and Sykhatme 2003-2006). Zufferey and Floreano mounted two cameras on a very light aircraft and developed optical flow based control algorithms for the aircraft flight within a room (Zufferey and Floreano 2006). Muratet et al. developed a navigation strategy for a virtual model of a helicopter that exploits optical flow information to fly between obstacles (Muratet et al. 2004).
Most of the previous works are about ground robots and emphasize the different methods of optical flow measurement. These works are usually experimental efforts without a clear discussion about the control and navigation algorithms. Experiments in computer vision researches are always susceptible to noise because of the variation in environmental lighting, vehicle vibration, sensor error, etc. and it is hard to distinguish between various parameters affecting the performance of a vision based navigation system. Therefore, a theoretical framework has been developed to model the optical flow measurement process using the camera model, the aircraft flight equations and the environment geometry. This framework gives the ideal optical flow information of each camera and makes it possible to study individually some of the parameters affect the navigation and control performance.
In this work, the balance strategy between the optical flow values of two cameras mounted on the aircraft is used to keep the aircraft in the middle of the obstacles. The environment is a corridor with parallel walls and a 3-way junction. The theoretical framework enabled us to study the effect of the ratio of the side way width to the main way width on the navigation performance of the aircraft. The effect of the camera field of view (FOV) angle on the navigation and control behavior are also discussed.

OPTICAL FLOW DEFINITION

Optical flow is the motion of picture when the camera moves relative to the environment. The mathematical relation between the optical flow and the camera status is as follows (Green et al. 2004):

$$OF = -\omega + \frac{v}{d}\sin\theta \qquad (1)$$

where OF is the one dimensional optical flow, ω and v are the camera rotational and translational velocities, d is the distance between the camera and the observed object, and θ is the object offset from the camera travel direction (Figure 1). As seen from the above equation, the OF vector can be described as a vector sum of two distinct components, one due to the translation, and another due to the rotation. The translation component is inversely proportional to the distance from objects. Therefore, if the rotational part can be eliminated, it would be possible to estimate the distances in each viewing directions.

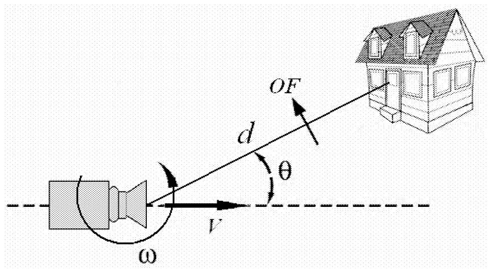


Figure 1. One dimensional optical flow

OPTICAL FLOW MODELING

Equation (1) is not practical for optical flow measurement and some image processing methods like gradient-based (Horn and Schunck 1981; Lucas and Kanade 1984) or correlation-based (Ancona and Poggio 1993) techniques are used in practice. In this work, however, we intend to model the measurement process of ideal optical flow using a camera model, the aircraft equations of motion and the environment geometry.

Coordinate Systems

Three coordinate systems that describe the geometrical relationships are the scene, the aircraft, and the camera coordinate systems. Figure 2 represents these coordinate systems. The scene coordinate is the inertial coordinate system with its origin fixed at the beginning and middle of the main corridor. The aircraft and camera coordinate systems are body coordinate systems that attach solidly on the aircraft c.g. and camera focal point, respectively.

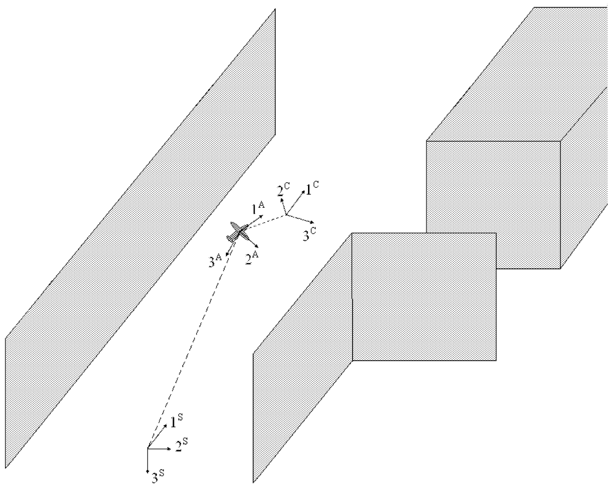


Figure 2. flight environment and defined coordinate systems

Transformation Matrices

It is assumed that the cameras are mounted on the aircraft such that the horizontal axis of the image plane, 1^C, lies on the $1^A 2^A$ plane of the aircraft. Therefore, the matrix that transforms a vector from the camera coordinate system to the aircraft coordinate system will be as follows:

$$[T]^{AC} = \begin{bmatrix} \sin\theta_c & \cos\theta_c \sin\varphi_c & \cos\theta_c \cos\varphi_c \\ \cos\theta_c & \sin\theta_c \sin\varphi_c & \sin\theta_c \cos\varphi_c \\ 0 & -\cos\varphi_c & \sin\varphi_c \end{bmatrix} \quad (2)$$

where θ_c and φ_c are camera installation angles. The transformation matrix between the aircraft and the scene coordinate systems is defined as following (Zipfel 2000):

$$[T]^{SA} = \begin{bmatrix} C\psi.C\theta & C\psi.S\theta.S\phi - S\psi.C\phi & C\psi.S\theta.C\phi + S\psi.S\phi \\ S\psi.C\theta & S\psi.S\theta.S\phi + C\psi.C\phi & S\psi.S\theta.C\phi - C\psi.S\phi \\ -S\theta & C\theta.S\phi & C\theta.C\phi \end{bmatrix} \quad (3)$$

where (ϕ, θ, ψ) are the aircraft Euler angles and S and C represent Sine and Cosine functions, respectively.

Camera Model

Here the Pinhole camera model adopted to simulate the camera measurement process. This model is a perspective projection model describes the mathematical relationship between the coordinates of a 3D point and its projection onto the image plane.

Geometrical Relationships

Figure 3 represents the geometrical relationship between the camera, the aircraft and the wall. The intersection point *P* can be calculated at time *t* having the focal point *O* and the pixel point *K* that already has been transformed to the scene coordinate system. These points are obtained using the following equations:

$$[r_O]^S = [r_A]^S + [T]^{SA}[AO]^A \quad (4)$$

$$[r_K]^S = [r_O]^S + [T]^{SA}[T]^{AC}[OK]^C \quad (5)$$

Vectors *AO* and *OK* in above equations are known in the aircraft and camera coordinates, respectively and are transformed to the scene coordinate system using the transformation matrices (2) and (3). Now, the point *P* can be calculated by the intersection of *OK* line and the wall plane defined in the scene coordinate system. For example, if the corridor walls are parallel to the 1^S axis, the walls plane equations in scene coordinate system will be as following:

$$(RW)_2^S = \frac{W}{2} \quad , \quad (LW)_2^S = -\frac{W}{2} \quad (6)$$

where *W* is the corridor width and *RW* and *LW* indicates the right and the left wall planes, respectively.

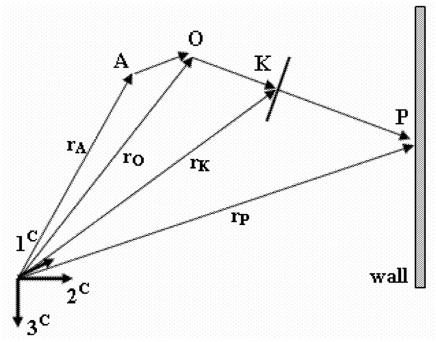


Figure 3 Geometrical relationships between the camera, the aircraft and the wall

The coordinate of the intersection point P in the scene coordinate system is obtained using the following equations:

$$P_1^S = O_1^S + \frac{K_1^S - O_1^S}{K_2^S - O_2^S}\left[(RW)_2^S - O_2^S\right]$$
$$P_2^S = (RW)_2^S \tag{7}$$
$$P_3^S = O_3^S + \frac{K_3^S - O_3^S}{K_2^S - O_2^S}\left[(RW)_2^S - O_2^S\right]$$

where $\left[P_1^S, P_2^S, P_3^S\right]^T$, $\left[K_1^S, K_2^S, K_3^S\right]^T$ and $\left[O_1^S, O_2^S, O_3^S\right]^T$ are the components of $[r_P]^S$, $[r_K]^S$, and $[r_O]^S$ vectors, respectively.

After a time step Δt, the position and orientation of the aircraft (and the cameras mounted on it) changes. By solving the flight equations, the new position of the aircraft $[r_{A'}]^S$ and the new transformation matrix $[T']^{SA}$ are calculated. The scene point P measured at time t now is connected to the new focal point that is available at time $t + \Delta t$ by the following equation:

$$[r_{O'}]^S = [r_{A'}]^S + [T']^{SA}[AO]^A \tag{8}$$

The line $O'P$ now is transformed to the camera coordinate system:

$$[r_{O'P}]^C = [T]^{CA}[T']^{AS}[r_{O'P}]^S \tag{9}$$

Therefore, the intersection of this line by the camera image plane can be calculated. Showing this intersection point by K', the optical flow vector will be as follows:

$$(OF) = \frac{O'K' - OK}{\Delta t} \tag{10}$$

This procedure is applied for all of the pixel points and the optical flow field is obtained for each of the cameras.

ELIMINATION OF ROTATIONAL OPTICAL FLOW

As the rate gyros are common sensors in aircraft and frequently are used in flight control applications, we assume that the rotational optical flow is eliminated having the aircraft rotational velocity measured by the rate gyros. The rotational optical flow can be obtained by the following equation:

$$(ROF)_i = -[OK_i]^C \times \left([T]^{CA}[\Omega]^A\right) \tag{11}$$

where $(ROF)_i$ is the rotational optical flow at the i^{th} pixel and Ω is the aircraft rotational velocity vector expressed in the aircraft coordinate system. The translational optical flow can be calculated simply by the vector subtraction.

AIRCRAFT CONTROL

Here, we used the balance strategy to control the aircraft flight path through the obstacles. In a classic turn, aircraft try to equalize the yaw angle ψ to a known value $\psi_{REF.}$ by changing the bank angle ϕ. The error signal in this controller is:

$$E = (\psi_{REF.} - \psi) \tag{12}$$

and with a simple proportional control, the aileron input value will be:

$$\delta_A = k_\phi\left[k_\psi(\psi_{REF.} - \psi) - \phi\right] \tag{13}$$

When an aircraft flies in a corridor, the approach angle is unknown and it is not possible to have a certain value for the reference yaw angle. Therefore, the error signal is defined as:

$$E = (OF_L - OF_R) \tag{14}$$

where OF_R and OF_L are the magnitudes of the optical flow field vectors in the right and left cameras, respectively. The following control law is used for the aircraft turn:

$$\delta_A = k_\phi\left[k_\psi \cdot (OF_L - OF_R) - \phi\right] \tag{15}$$

This is the strategy that some of the flying insects use to decide the turn direction (Tammero and Dickinson 2002).

SIMULATION AND RESULTS

As shown in figure 2, the environment is a main corridor with two parallel walls and a 3-way junction. It is assumed that the walls are high enough that the cameras don't see the ground or the sky while the pitch angle varies. It is also assumed that the main corridor has 40m width and the aircraft flies through it with an initial speed of 5 m/s. The specifications of the cameras was selected based on the existing small cameras in the market.

Aircraft Dynamics

The aircraft dynamics is modeled by the six degree of freedom flight equations (Zipfel 2000):

$$\dot{u} = rv - qw - g\sin\theta + \frac{F_x}{m}$$
$$\dot{v} = -ru + pw + g\sin\phi\cos\theta + \frac{F_y}{m} \tag{16}$$
$$\dot{w} = qu - pv + g\cos\phi\cos\theta + \frac{F_z}{m}$$

$$\dot{p} = (C_1 r + C_2 p)q + C_3 L + C_4 N$$
$$\dot{q} = C_5 pr - C_6 (p^2 - r^2) + C_7 M \quad (17)$$
$$\dot{r} = (C_8 p - C_2 r)q + C_4 L + C_9 N$$

where (u, v, w) and (p, q, r) are the components of the aircraft translational and rotational velocity, m is the aircraft mass, g is the gravitational acceleration, $C_{1,2,\ldots,9}$ are the coefficients of the aircraft moments of inertia, and (F_x, F_y, F_z) and (L, M, N) are the components of the aerodynamic and thrust forces and moments exerted on the aircraft. In this work, the aerodynamic force and moments are modeled using constant stability derivatives (Roskam 1998). Six more equations needed to obtain the aircraft position and orientation. These equations are the three kinematic and three navigation equations:

$$\dot{\phi} = p + \tan\theta(q \sin\phi + r\cos\phi)$$
$$\dot{\theta} = q\cos\phi - r\sin\phi \quad (18)$$
$$\dot{\psi} = (q\sin\phi + r\cos\phi)/\cos\theta$$

$$\dot{\mathbf{U}} = [T]^{SA}\mathbf{V} \quad (19)$$

where $\mathbf{U} = [U_1, U_2, U_3]^T$ is the aircraft position vector in the scene coordinate system and $\mathbf{V} = [u, v, w]^T$ is the aircraft velocity vector.
The aircraft used in this work is a very light aircraft and has a weight of only 90 grams. The aircraft specifications are presented in table 1 based on the (Foster et al. 2005). The aircraft mass properties were changed to obtain lower flight speeds in trim condition.

Table 1: Aircraft geometry and mass properties

Mass	m = 0.091 *(kg)*
Moment of inertia	I_{11} = 0.0004 *(kg.m²)*
Moment of inertia	I_{22} = 0.0008 *(kg.m²)*
Moment of inertia	I_{33} = 0.001 *(kg.m²)*
Wing area	S = 0.089 *(m²)*
Wing span	b = 0.61 *(m)*

Aircraft Navigation Through Corridor

At the beginning of the simulation, it is assumed that the aircraft is initially at the origin of the scene coordinate system and flies parallel to the walls in the main way. It is also assumed that the cameras' horizontal and vertical FOV angles are 42° and 10°, respectively.
The first simulation was carried out with the ratio of the side way width to the main way width equals to 0.5. Figure 4 shows the aircraft horizontal flight path. The aircraft begins to deviate toward the side way approximately 20m before the junction. This is the point that the right camera sees the side way for the first time. The aircraft deviated about 4m from the initial direction but returned back to it after passed the junction.
Figure 5 shows the aircraft flight path for the larger widths of the side way. It is seen that when the widths are equal (w_s/w_m=1) the aircraft deviates into the side way but still is unable to enter it and returns back to the main way in the reverse direction. However, when the width of the side way is increased slightly (w_s/w_m=1.05), the aircraft enters the side way and tries to keep itself in the middle of the new corridor.

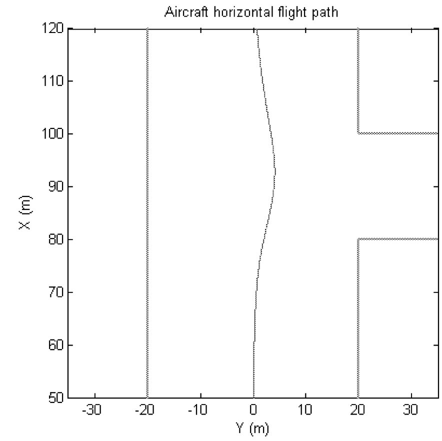


Figure 4. Aircraft flight path in width ratio of 0.5

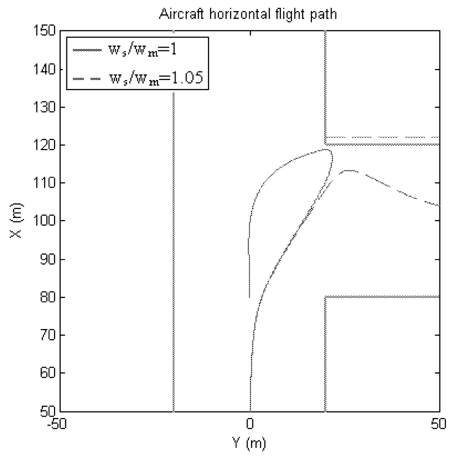


Figure 5. Aircraft horizontal flight path at the width ratios of 1 and 1.05

Effect of the FOV Angle of the Cameras

Another parameter that may affect the navigation performance is the FOV angle of the cameras mounted on the aircraft. Figure 6 shows the aircraft flight path at the width ratio of 1.05 for the three different horizontal FOV angles. It is seen that the larger FOV angle improves the navigation performance and the aircraft is settled more rapidly on the center line of the side way. However, when the FOV angle is small, the aircraft couldn't see the side way well and hence is unable to enter it.

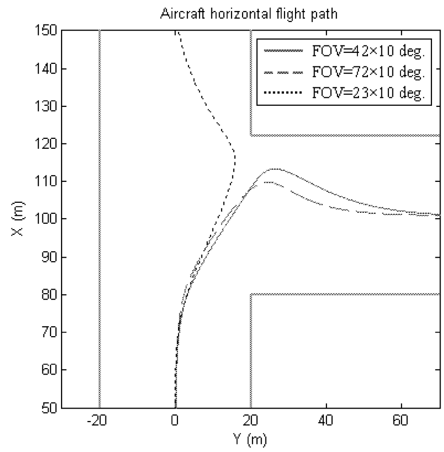


Figure 13. Effect of the cameras' horizontal FOV angle on the navigation performance at the width ratio of 1.05

CONCLUSION AND FUTURE WORK

The optical flow information of two cameras mounted on both sides of an altra light aircraft was simulated using a geometrical framework consists of perspective camera model, aircraft equations of motion and the geometry of environment. This information then was used to navigate the aircraft in a 3-way junction using the balance strategy.
The simulations showed that when the width of the side way is bigger than the main way, the aircraft is deviated completely toward the junction and enters into the side way. The simulations also showed that as the FOV angle of the cameras increases, the navigation performance and the aircraft motion behavior at the junction is improved.
The proposed framework gives ideal optical flow filed based on the geometrical relationships. In reality, however, the measured optical flow is affected by various parameters like image distortion, lighting levels, texture variation, etc. In the future we hope to complete the proposed method using more exact camera models, inclusion of the environment texture and lighting condition, and optical flow measurement by conventional image processing methods.

REFERENCES

Ancona, N. & Poggio, T. 1993. "Optical flow from 1d correlation: Application to simple time-to-crash detector ", *International Journal of Computer Vision*, Vol. 14, No. 2, 131-146.

Argyros, A. & Bergholm, F. 1999. "Combining central and peripheral vision for reactive robot navigation", *In Proc. Computer Vision and Pattern Recognition Conference (CVPR'99)*, Fort Collins, Colorado, USA, 2646–2651.

Argyros, A. Tsakiris, D.P. and Groyer, C´. 2004. "Biomimetic centering behavior for mobile robots with panoramic sensors", *IEEE Robotics and Automation Magazine*, Vol. 11 No. 4, 21–68.

Barrows, G. L. 1999. "*Mixed-Mode VLSI Optic Flow Sensors for Micro Air Vehicles.*", Ph.D. Dissertation, Department of Electrical Engineering, University of Maryland at College Park, College Park, MD.

Camus T, Coombs D, Herman M, and Hong T. H. 1999, "Real-time single workstation obstacle avoidance using only wide-field flow divergence", *Journal of Computer Vision Research*, Vol. 1, No. 3, 30–57.

Foster, T. M. & Bowman, W. J. 2005. "Dynamic stability and handling qualities of small Unmanned-Aerial-Vehicles", *43rd AIAA Aerospace Sciences Meeting and Exhibit*, Reno, Nevada.

Green W. E, Oh P. Y, Sevcik K, and Barrows G. 2003. "Autonomous Landing for Indoor Flying Robots Using Optic Flow", *Proceedings of ASME International Mechanical Engineering Congress and Systems (IMECE)*, Washington, DC, 1347-1352.

Green W. E, Oh P. Y. and Barrows G. L. (2004), "Flying insect inspired vision for autonomous aerial robot maneuvers in near-earth environments", *In Proc. IEEE International Conference on Robotics and Automation*, Vol. 3, New Orleans, LA, 2347–2352.

Horn B.K.P. and Schunck B.G. 1981. "Determining optical flow", *Artificial Intelligence*, Vol. 17, 185-203.

Hrabar, S. & Sukhatme, G. S. 2003. "Omnidirectional Vision for an Autonomous Helicopter", *IEEE International Conference on Robotics and Automation*, 558 - 563.

Hrabar S., and Sukhatme G. 2004. "A Comparison of Two Camera Configurations For Optic-Flow Based Navigation of a UAV Through Urban Canyons", *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2673 - 2680.

Hrabar S, Sukhatme G, Corke B, Usher K. and Roberts, J. 2005. "Combined Optic-Flow and Stereo-Based Navigation of Urban Canyons for a UAV", *Intelligent Robots and System, (IROS 2005). 2005 IEEE/RSJ International Conference on*, 3309-3316.

Hrabar, S. & Sukhatme, G. S. 2006. "Optimum Camera Angle for Optic Flow-Based Centering Response", *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, 3922-3927.

Lucas, B. D. and Kanade, T. 1981. "An iterative image registration technique with an application to stereo vision", *Proceedings of Imaging understanding workshop*, 121-130.

Muratet A, Doncieux S, and Meyer J. A. 2004. "A biomimetic reactive navigation system using the optical flow for a rotary-wing UAV in urban environment", *Proceedings of the International Session on Robotics*, 2004.

Oh P.Y, Green W. E. and Barrows G. 2004. "Neural Nets and Optic Flow for Autonomous Micro-Air-Vehicle Navigation", *Proceedings of 2004 ASME International Mechanical Engineering Congress and Exposition (IMECE)*, 2004, Volume 2, Anaheim, California USA, 1-7.

Roskam, J. 1998. *Airplane Flight Dynamics and Automatic Flight Controls, Part I,* DARcorporation, Kansas, USA.

Tammero, L. F. & Dickinson, M. H. 2002. "The influence of visual landscape on the free flight behavior of the fruit fly drosophila melanogaster", The Journal of Experimental Biology, Vol. 205, 327-343.

Victor, J. S. & Sandini, G. 1995. "Divergent stereo in autonomous navigation: From bees to robots", *International Journal of Computer Vision*, Vol. 14, No. 2, 159-77.

Zipfel, P. H. 2000. *Modeling and simulation of aerospace vehicle dynamics*, AIAA education series.

Zufferey, J. C. 2005. "*Bio-inspired Vision-based Flying Robots*", PhD thesis, Swiss Federal Institute of Technology in Lausanne (EPFL).

Zufferey J. C. & Floreano, D. 2006. "Fly-inspired visual steering of an ultralight indoor aircraft", *Robotics, IEEE Transactions on*, Vol. 22, 137-146.

Towards Fine-grained Spatial Partition for Wildfire Simulation

Song Guo, Xiaolin Hu
Department of Computer Science
Georgia State University
34 Peachtree Street, Suite 1450,
Atlanta, GA, 30303, USA
Guosong1079@hotmail.com, xhu@cs.gsu.edu

KEYWORDS
Performance, DEVS-FIRE, Spatial Partition, PDES (Parallel Discrete Event Simulation), the Four Color Theorem

ABSTRACT

Tasks partitioning is one of the most challenging issues in large scale parallel simulations. The simulation performance may vary dramatically if different partition schemes are used. This paper presents a spatial partition method named fine-grained spatial partition for parallel simulation of large scale wildfire. The fine-grained spatial partition is inspired by the four color theorem. It divides a cell space into many smaller regions and allocates adjacent regions to different processing units. Because no adjacent regions are on the same processing units, a fire started in one region will quickly spreads out to adjacent regions and thus achieves a higher degree of parallelism regardless where the fire is started. Experiment results are collected and compared with that of a uniform spatial partition.

1. INTRODUCTION

The startling damage brought by wildfires has drawn more and more attention from all over the world during past decades. The estimated loss of the 2008 California Wildfire is about 1.4 billion dollars. It is important for us to study the behavior of the fire spread process and to come up with strategies to control, fight against and suppress the fire. Many simulation tools, such as FARSITE (Finney 1998), BehavePlus (Andrews et al. 2005), Hfire (Morais 2001), and DEVS-FIRE (Natimo et al. 2008), are developed with the purpose of providing comprehensive and accurate wildfire simulation.

In the previous work, we have developed a wildfire simulation model called DEVS-FIRE. DEVS-FIRE is built on the DEVS formalism (Zeigler et al. 2000). It employs an M*N two-dimensional cellular space as the target area where the fire occurs. The cells in the space are used to represent sub regions of the whole area and each cell is an atomic model in DEVS. For large-scale simulation of DEVS-FIRE, parallel simulation is essential. One of the challenging problems in parallel simulation is how to partition the simulation tasks among multiple processing units (PUs). In (Guo et al. 2009), we explored several ways of spatial partition, among which a uniform partition divides the whole space into several size-equivalent spaces and assigns each PU one of these spaces. The deficiency of this approach is that the locations and the number of the ignition points may affect the performance of the parallel simulation. The usage of the PUs is also affected by those two factors.

In this paper, we propose a new partition approach inspired by the four color theorem to improve the performance of parallel simulation of DEVS-FIRE. In the new method, we divide the entire cellular space into a number of much smaller, size-equivalent cellular spaces. Borrowing from the four color theorem, we regard these smaller cellular spaces as **region**s that need to be colored on a map. The number of processing units is regarded as the number of colors we can use. Therefore, the process of partitioning the simulation tasks can be viewed as the process of coloring the map. The goal is to achieve a result that any regions that neighbor with each other will have different colors, so the adjacent regions will be allocated to different PUs. Since the size of each region is not very large, the fire started in one region will quickly spreads out to adjacent regions at an early stage of the parallel simulation. This spread-out will trigger other PUs to run. In this way, we aim to achieve a higher degree of parallelism no matter where the locations of the ignition points are. The design of the new method is presented in this paper. The experiment results are compared with that of the uniform spatial partition approach.

The remainder of the paper is organized as follows. Section 2 briefly reviews related applications of PDES and load-balancing strategies. Section 3 gives an overview of the uniform spatial partition for parallel simulation of DEVS-FIRE. Section 4 presents the basic idea of our fine-grained partition for the parallel simulation of DEVS-FIRE. Two possible partition algorithms are also provided. Section 5 presents the experiments design and shows the experiment results and analysis. Section 6 draws conclusions and points out future work.

2. RELATED WORK

During the recent decades, parallel simulation has drawn great attention of people from all over the world as the parallel and distributed computing develops. Task partitioning plays a significant role in the performance of parallel simulation. The partitioning is known to be NP-complete in general (Garey et al. 1976). However, when the underlying task exhibits certain regularities, good partitions can be found in polynomial time. In the work of (Glöckel and Fleischmann 2000), the authors present a new approach for partitioning VLSI circuits on transistor level and gives runtimes of parallel simulations of large industrial circuits. Their scheme is based on a clustering algorithm with a new signal model concept and a special coupling measure. This enables a very efficient cluster growth. Their experimental results show considerable improvement to a known partitioning method, the Node Tearing method. In (Kim et al. 1997), Kim and Park proposed a concurrency preserving partitioning algorithm for the optimistic parallel simulation of hierarchical, modular discrete event models. Their novel partitioning algorithm utilized the hierarchical structural information of models with the aim to balance the computational loads of partitions, to maximize the parallel execution of independent models and to minimize the inter-processor communications. Partitioning and parallel simulation is widely used in the area of network as well. The work of (Ammar and Deng 1991) gave an approach of parallel simulation to divide a general stochastic Petri net spatially into several connected subnets. Their work can also be used to handle general forms of network partitions.

3. SPATIAL PARTITION FOR PARALLEL SIMULATION OF DEVS-FIRE

DEVS-FIRE is an integrated simulation environment for surface wildfire spread and containment built on the Discrete Event System Specification (DEVS) formalism. In DEVS-FIRE, a forest is modeled as a two dimensional cellular space composed of individual forest cells coupled together according to their relative physical geometric locations. Each cell represents a sub area of the target forest and each cell is coupled to its eight neighbors corresponding to the N, NE, E, SE, S, SW, W, and NW directions respectively. Accordingly, fire spreading is modeled as a propagation process as burning cells ignite their unburned neighbor cells. More information about the DEVS-FIRE model can be found in (Natimo et al. 2008).

We explored a uniform spatial partition approach in our previous paper (Guo et al. 2009). The basic idea of this approach is to divide the two-dimensional cellular space of DEVS-FIRE into multiple smaller size-equivalent cellular spaces and allocate each partition to a PU for parallel simulation. Simple as it is, this partition method suffers two defects. One defect is that the number and the locations of the ignition points can significantly influence the performance of parallel simulation. For example, if we have a 400×400 cellular space and 4 processing units, we can partition this cellular space into 4 200×200 cellular spaces. However, if there is only one ignition point, the performance will vary a lot as the location of the ignition point moves from corner to the partition boundary (Guo et al. 2009). The utilization of each PU also varies as the ignition point moves. If the ignition point locates far away from the partition boundary, only one processing unit is working while the other three are doing nothing but waiting for the burning messages. The other defect is that the number of PUs involved in the parallel simulation is somehow restricted by how the space is divided. Generally, it is better to use square number (such as 4, 9, 16, …, N^2, where N is an integer) of PUs under uniform partition method. The reason is that we can result a hybrid partitioning rather than a partitioning in horizontal or vertical fashion. We note that it is still possible to produce a partitioning in a horizontal fashion or in a vertical fashion (see Figure 1 as an example) using square number of PUs. However, the result is either unbalanced or has poor parallelism. The advantage of hybrid partitioning is that each partition has more neighbor partitions (at least 3 neighbor partitions and 8 at most) than in horizontal or vertical partitioning (at most 2). Figure 1 shows a horizontal partitioning and a hybrid partitioning of a cell space using 9 PUs. In this example, if the fire is ignited at the top right corner of the partition0, the PUs in horizontal partitioning will be poorly utilized during the run time of the parallel simulation because the PUs work nearly sequentially. However, in hybrid partitioning, fire can spread from partition0 to partition1, partition3 and partition4, thus achieving more parallelism than in horizontal partitioning.

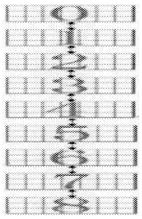


Figure 1(a): A horizontal uniform partition

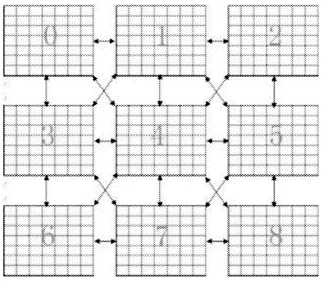


Figure 1(b): A hybrid uniform partition

Figure 1: uniform partitioning using 9 PUs

4. MAP COLORING-INSPIRED PARTITION FOR PARALLEL SIMULATION OF DEVS-FIRE

Partition benefits the performance of the parallel discrete event simulation, but the uniform spatial partition approach has some limitations. Motivated by this, we propose a new approach which is inspired by the map coloring problem. The main idea of the map coloring-inspired partition approach is to spatially divide the cellular space into a number of smaller cellular spaces and distribute those smaller cellular spaces among multiple processing units so that no two adjacent spaces are on the same processing unit.

4.1. Map Coloring

In mathematics, there is a famous theorem, namely the **four color theorem** or the **four color map theorem**, stating that given any separation of a plane into contiguous regions, called a *map*, the regions can be colored using at most four colors so that no two adjacent regions have the same color. Two regions are called *adjacent* only if they share a border segment, not just a point (Georges 2008). Five colors are suffices to color a map; however using four colors turned out to be significantly harder. The four color theorem has not been truly proven until 1976 under the help of the computer. Although the four color theorem was proven, the theorem itself does not give any algorithm about how to color. Figure 2 illustrates an example of four color theorem.

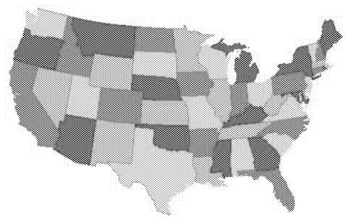


Figure 2: A four-coloring of the map of the states of the United States (except Alaska and Hawaii)

4.2. Map Coloring-inspired Partition

Constructed by thousands of cells, the two-dimensional cellular space of DEVS-FIRE can be viewed as a map. To make the partitions smaller, either one cell or a group of cells can be viewed as the smallest region to color. The "group" here means that those cells construct a rectangle or a square. By using these two shapes, the partition procedure could be much easier. Other shapes, such as triangles, surely work; but it would be difficult to implement. As for our case, the definition of "adjacent" is a little stricter in the parallel simulation of DEVS-FIRE because each cell has eight neighbors and it can communicate with all its eight neighbors. This makes each region have eight neighbors. So two regions are adjacent if they either share a border segment, or they only share a point. In our approach, the number of participating PUs can be viewed as the number of available colors. Therefore, the process of partitioning can be viewed as the process of coloring: try to "color" two adjacent regions using two different "colors".

However, there exists a difference between our case and map coloring. To achieve higher performance of our parallel simulation of DEVS-FIRE, all participating PUs are put into use and the number of regions each PU obtains should be equal or close. While four or five colors are adequate to color a map. The extra colors can be used but not necessary. Moreover, one does not need to guarantee whether each color is equally used. Therefore our partition approaches do not necessarily stick to the four color theorem. The main goal is to scatter the fine-grained regions among available PUs.
Before we present the new partition algorithm, some parameters must be specified. For simplicity, we only use M×M cellular space for parallel simulation of DEVS-FIRE. N is the number of participating PUs. Each PU is represented by PU_i, where $0<=i<N$. Parameter g is used to denote that a $g \times g$ square is "colored" each time (a rectangle works too, but we use a square in the implementation for simplicity. See Figure 3 as an example). We refer this $g \times g$ square as **region** in the rest of this paper. Cell(x,y) represents the

cell locates at (x, y) of the cell space. After "coloring", each PU will get a set of cells, which is represented as $\mathbf{S}_i$.

Next we present two possible partition algorithms that follow our goal.

(1) First Situation: $M/g >= N$

```
low_w := M / g
low_h := M / g
for xl := 0 to low_w-1
   for yl := (N-xl+i+N*low_h)%N to low_h-1
      high_w_start := xl * g
      high_w_end := high_w_start + g
      high_h_start := yl * g
      high_h_end := high_h_start + g
      for x := high_w_start to high_w_end-1
         for y := high_h_start to high_h_end-1
            Cell(x, y) ∈ S_i
```

Deducted from the algorithm above, we have the cell set that PU_i gets:

$\mathbf{S}_i = \{Cell(x, y) \mid xl*g<=x<xl*g+g, yl*g<=y<yl*g+g, 0<=xl, yl<M/g, yl= (N-xl+i+N*M/g) \%N\}$.

Figure 3 shows two partition examples of the new algorithm when $M/g >= N$.

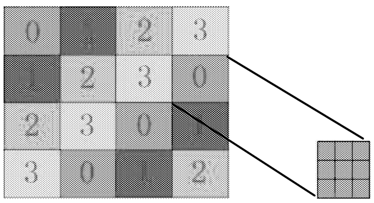


Figure 3(a): an example of the algorithm when M=12, N=4, g=3

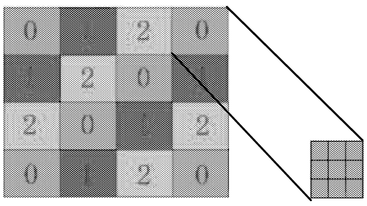


Figure 3(b): an example of the algorithm when M=12, N=3, g=3

Figure 3 two partition examples when $M/g >= N$

(2)Second Situation: $M/g < N$

```
low_w := M / g
low_h := M / g
t := 0
count = 2 * low_w - 1
for z := 0 to count-1
   for xl := 0 to low_w-1
      for yl := low_h-1 to 0
         if xl+yl == z
            if t == i
               high_w_start := xl * g
               high_w_end := high_w_start + g
               high_h_start := yl * g
               high_h_end := high_h_start + g
               for x := high_w_start to high_w_end-1
                  for y := high_h_start to high_h_end-1
                     Cell(x, y) ∈ S_i
            End if
            t++
            t % = N
         End if
```

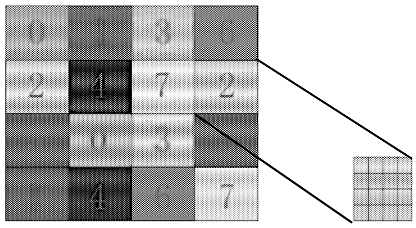


Figure 4: an example of the algorithm when M=16, N=8, g=4 (16/4<8)

The second part ($M/g < N$) works as a remedy for the first one. The basic idea of the second part is to achieve load balance and make adjacent regions on different PUs. For the first part of the new algorithm, we can accept the partition when N is a little larger than M/g (as shown in Figure 5). The problem with this situation is that the load is not well balanced. When N is much larger than M/g, say N = 16; then some PUs will not even secure a region at all according to the first part of the new algorithm. However, the size of the cellular space is usually much larger than the number of PUs and we can always choose a good relationship between M/g and N to avoid this defect in the parallel simulation.

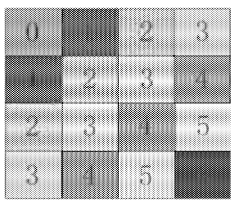


Figure 5: an example using the first part of the algorithm when $M/g = 4$, $N = 7$

5. EXPERIMENT RESULTS

In this section, we designed several experiments to show the simulation results of our algorithm. The simulations were conducted on Cheetah, a Linux cluster server, which contains 5 nodes and 4 GPUs. The software we used is adevs-2.1 and mvapich2/gnu. In all the experiment, we assume that the wind direction and wind speed remain constant while doing the simulation. We use uniform terrain map for all our experiments and the simulations finished when all the cells in the cellular space are burned.

5.1. Experiment 1: Compare the Uniform Spatial Partition with the Map Coloring-inspired Partition

This experiment is conducted to test the execution time and the total number of rollbacks of both approaches on different sizes of cellular spaces. The sizes of cellular spaces used are 400×400, 600×600, 800×800, and 1000×1000, one for each round. In this experiment, we use one ignition point and four processing units. The location of this ignition point is at (0, 0). The location of (0, 0) is located at the bottom left corner of the cell space. The parameter g is set to 20.

Figure 6 displays the execution time and the total number of rollbacks for each round. From the experiment results, we can see that our map coloring-inspired partition has better performance than the uniform spatial partition in execution time. However, uniform partition has less rollback. This is because that most of the time, the computation is local. Rollbacks happen when the fire spread out of the current partition. At that time, we need to send burning messages to other PUs and that will cause the rollbacks. Moreover, the work amount graphs, which represent the number of iterations in the simulator during a certain time, indicate that our new approach utilizes PUs better: the working period of each PU overlaps from the beginning to the end of the simulation. This is true because no matter wherever the ignition point moves, the fire will spread out to multiple regions, which are on different PUs. This will make multiple PUs start to run at an early stage of the parallel simulation if the g is not very large.

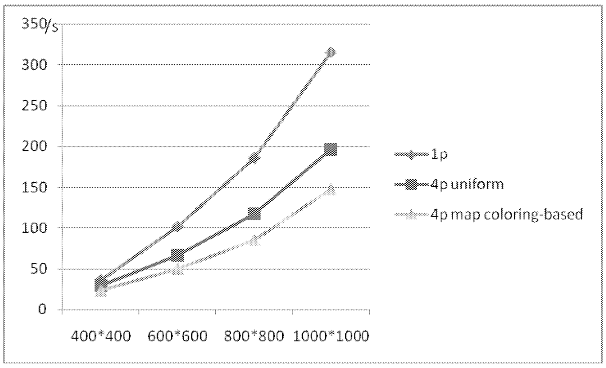


Figure 6(a): execution time of different sizes of the cellular spaces with four processors and one ignition point at (0, 0)

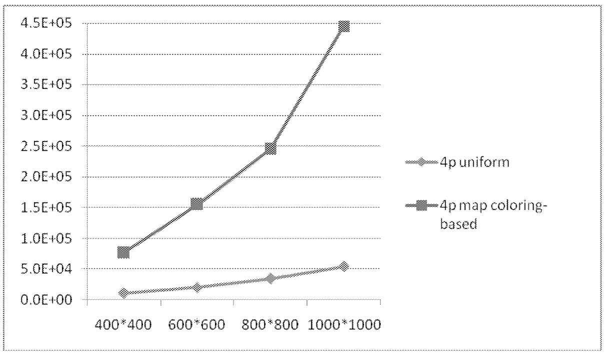


Figure 6(b): the number of rollbacks

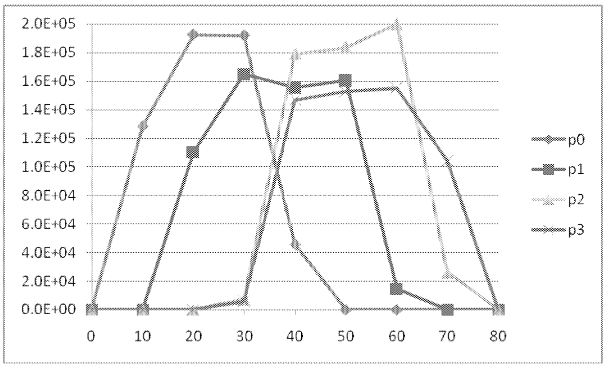


Figure 6(c): work amount at run time of uniform partition using 600×600 cell space

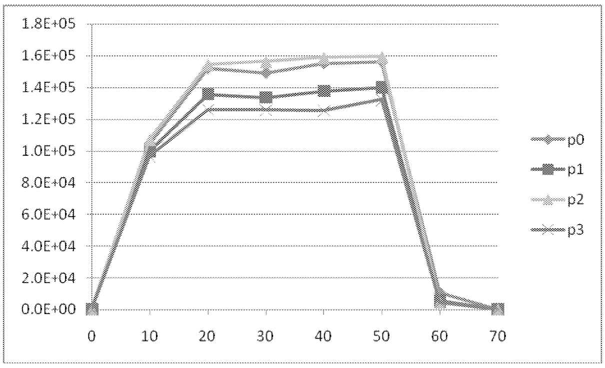


Figure 6(d): work amount at run time of map coloring-inspired partition using 600×600 cell space

5.2. Experiment 2: Locations of the Ignition points vs. Performance

This experiment takes the locations of the ignition points into consideration and tests the effects this factor plays on the performance of the parallel simulation of both approaches. Four ignition points are set at different positions in a 400×400 cellular space. Table 1 gives the detailed locations of the ignition points. The parameter *g* is set to 20. Four processing units are employed in that we can result in a good partition for the uniform spatial partition.

Table 1: Ignition Points Settings

round	ignition points locations
1	(99,99),(299, 99), (99,299), (299,299)
2	(0,0), (399,0), (0,399), (399,399)
3	(199,199),(199,200),(200,199),(200,200)

Figure 7 shows the results of the this experiment, which illustrates that the locations of the ignition points nearly have no effect on our new approach while the performance varies a lot for the uniform spatial partition.

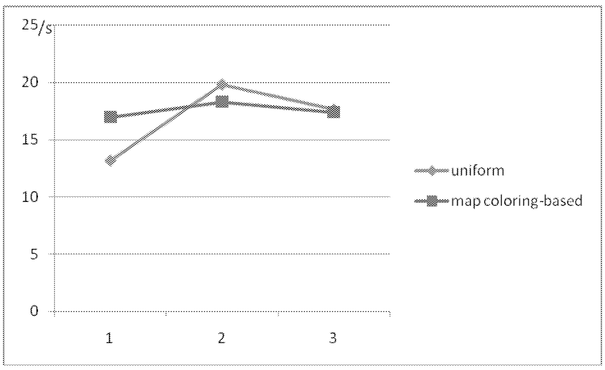


Figure 7(a): execution time

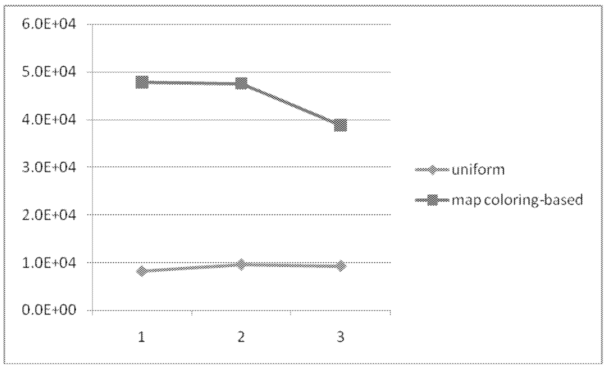


Figure 7(b): the number of rollbacks

5.3. Number of PUs vs. Execution Time in Map Coloring-inspired Partition

This experiment varies the number of the PUs on a 400×400 cellular space to see the scalability of the map coloring-inspired partition. The ignition point locates at (199,199) and *g* is set to 20.

Figure 8 displays the result of using different number of PUs of our new method. As the number of PUs increases, the execution time decreases. It also displays that our approach has no strict requirement for the number of PUs while only square number PUs may result a good partition for the uniform spatial partition approach.

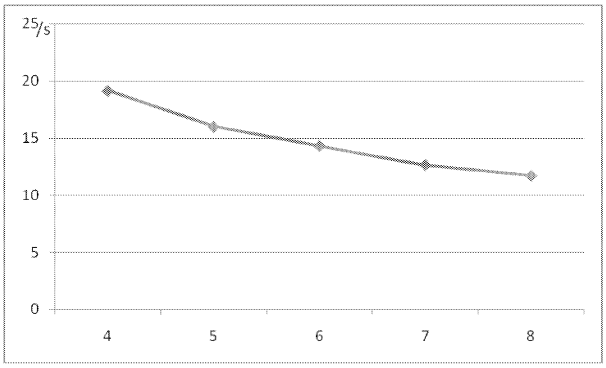


Figure 8(a): execution time

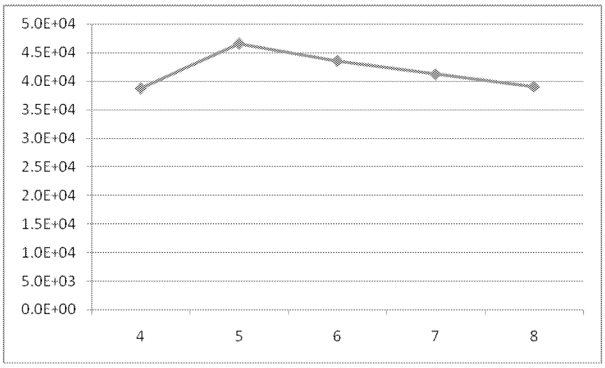


Figure 8(a): number of rollbacks

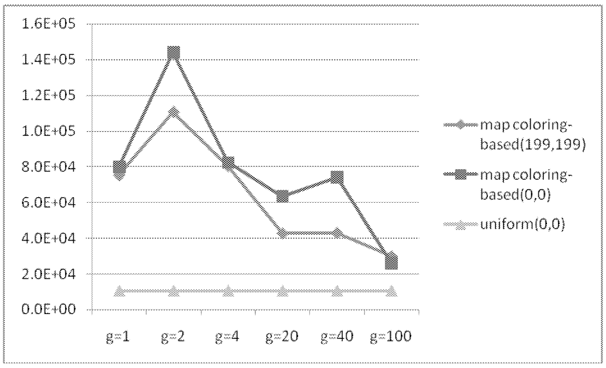


Figure 9(b): number of rollbacks with different values of parameter *g*

5.4. The Effect of Parameter *g*

For the last experiment, we set *g* to different values to test the effect it plays on the simulation. The size of the cellular space used in this experiment is 400×400 and four PUs are employed. Only one ignition point is used and we set it at two different locations, (0, 0) and (199, 199) respectively. As a comparison, we borrow the experiment results of the uniform spatial partition on the same 400×400 cellular space from experiment 1.

Both the figures in Figure 9 indicate that the parameter *g* can affect the performance of the new method. A good *g* can result in a good performance. Too small and too large represent two extremities. Too small, the communication or message passing will dominate; while if too large, the partition will be much closer to the uniform spatial partition.

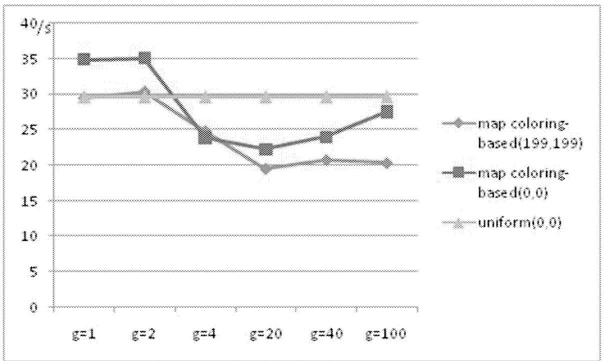


Figure 9(a): execution time with different values of parameter *g*

6. CONCLUSIONS AND FUTURE WORK

We proposed a new partition method which is inspired by the map coloring problem. Two practical partition algorithms are provided. Preliminary experiment results show that the partition method is not only stable but also overcomes the deficiencies of the uniform spatial partition in most situations.

Based on the preliminary experiment results, our future work will focus on the following directions: 1) formal performance analysis of the map coloring-inspired partition method; 2) combine real GIS and weather data into our parallel simulation of DEVS-FIRE; and 3) study the applicability of this partition method for other types of applications.

References

Andrews, Patricia L., Bevins, Collin D., Seli, Robert C, 2005, "BehavePlus fire modeling system, version 4.0: User's Guide", Gen. Tech. Rep. RMRS-GTR-106WWW Revised. Ogden, UT: Department of Agriculture, Forest Service, Rocky Mountain Research Station. 132p.

Gonthier, Georges, 2008, "Formal Proof—The Four-Color Theorem", *Notices of the American Mathematical Society* **55**(11).

H.H. Ammar and S. Deng, 1991, "Parallel Simulation of Stochastic Petri Nets Using Spatial Decomposition", Invited Paper, *Proc. of the 1991 International Symposium on Circuits and Systems, Singapore*.

Ki Hyung Kim, Tag Con Kim, Kyu Ho Park, 1997, "A concurrency preserving partitioning algorithm for parallel simulation of hierarchical, modular discrete event models", hpcasia, pp.204, *High-Performance Computing on the Information Superhighway, HPC-Asia '97*.

L. Ntaimo, X. Hu, and Y. Sun, 2008, "DEVS-FIRE: Towards an Integrated Simulation Environment for Surface Wildfire Spread and Containment", *SIMULATION*, Vol. 84, Issue 4, pp. 137-155.

Mark A. Finney, 1998, "FARSITE: Fire Area Simulator–Model Development and Evaluation", United States

Department of Agriculture Forest Service Rocky Mountain Research Station Research Paper, RMRS-RP-4 Revised March 1998, revised February 2004.
Morais, M., 2001, "Comparing Spatially Explicit Models of Fire Spread through Chaparral Fuels: A New Model Based upon the Rothermel Fire Spread Equation", *MA Thesis*, University of California, Santa Barbara.
M. R. Garey, D. S. Johnson and L. Stockmeyer, 1976, "Some Simplified NP-Complete Graph Problems", *Theoretical Comp. Sci.*, 1(1976), 237-267.
N. Fröhlich, V. Glöckel and J. Fleischmann, 2000, "A New Partitioning Method for Parallel Simulation of VLSI Circuits on Transistor Level", *Proc. of the Conference on Design, Automation and Test in Europe*, Paris, France, pp 679-685.
S. Guo, X. Hu, Y. Sun, 2009, "Exploring Spatial Partition for Parallel Simulation of DEVSFIRE", *Proc. Summer Computer Simulation Conference (SCSC'09)*.
Zeigler, B. P., 2000, H. Praehofer, and T. G. Kim, Theory of modeling and simulation, 2nd Edition, *Academic Press*.

TRANSPORT SIMULATION

STRATEGIC TRANSPORT DECISION-MAKING: THE SIMDEC APPROACH BASED ON RISK SIMULATION AND MULTI-CRITERIA ANALYSIS

Steen Leleur, Luise Augusta Larsen & Britt Zoëga Skougaard
Decision Modelling Group, DTU Transport
Technical University of Denmark (DTU)
DK 2800 - Lyngby, Denmark
sl@transport.dtu.dk

KEYWORDS
Transportation, decision-making, computer-aided analysis

ABSTRACT

This paper presents a new approach, SIMDEC, based on using risk simulation and multi-criteria analysis in combination for decision support. SIMDEC aims at providing decision support for transport decision-making with a focus on awareness of feasibility risks and inclusion of non-monetary, strategic concerns. First the SIMDEC modelling framework and the theories behind are described, and afterwards SIMDEC is illustrated by an example concerning examination of four infrastructure alternatives between Helsingør (Elsinore) and Helsingborg. These alternatives need to be assessed to facilitate decision-making concerning a second northern fixed link between Denmark and Sweden to supplement the already established fixed link in the southern part of Øresund connecting Copenhagen and Malmø. Finally, after a discussion of results, a conclusion is given together with a perspective on the further application of SIMDEC.

INTRODUCTION

Providing suitable decision support for strategic transport decision-making is a topic of growing concern. For large infrastructure investments, to exemplify with one important transport topic area, comprehensive assessments are needed (Banister & Berechman, 2000). Typically such investments have many-sided consequences which ought all to be taken into consideration to seek out the best alternative from a set of candidates that has come forward from the preparatory planning and design phases. The traditional cost-benefit analysis (CBA) as prescribed in various national and international appraisal manuals is insufficient as often important decision factors such as environment, regional development, etc. are not possible to cover by CBA (Leleur, 2000). Furthermore, issues about uncertainty take on a major role in connection with large scale projects, where factors such as construction costs and demand prognoses are uncertain for a number of reasons but clearly of very high importance for the long-term feasibility of the investment. Many cases have been documented where uncertainty of these factors have led to investments that later on turned out to be less than satisfactory (Priemus et al., 2008).

This paper presents a new approach to strategic transport decision-making, SIMDEC, based on using risk simulation and multi-criteria analysis in combination for decision support. First the SIMDEC modelling framework and the theories behind are described, and afterwards SIMDEC is illustrated by an example concerning examination of four alternatives between Helsingør (Elsinore)-Helsingborg. These alternatives need to be assessed to facilitate decision-making about a second northern fixed link between Denmark and Sweden to supplement the already established fixed link in the southern part of Øresund between Copenhagen and Malmø. Finally, after a discussion of results, a conclusion is given together with a perspective on the further application of SIMDEC.

THE SIMDEC MODELLING FRAMEWORK

As mentioned in the introduction, SIMDEC is based on applying risk analysis in combination with multi-criteria analysis. With the focus of the risk analysis on the feasibility of each of the alternatives, this concerns feasibility risk assessment (FRA) where the scope is on the risk that the investment could turn out not to be socio-economically feasible. SIMDEC proceeds by first examining FRA for the alternatives one by one, and afterwards the FRA results are used as input as one of the criteria within a set of decision criteria to a multi-criteria analysis (MCA) that aims at ranking the alternatives. The two major components of the SIMDEC modelling framework are described below.

Feasibility risk assessment (FRA)

The FRA is carried out by using Monte Carlo simulation (Vose, 2002) (Salling, 2008) on the results stemming from a conventional cost-benefit analysis, which is assumed to be prescribed by and conducted in accordance with a manual that can generally be accepted in the study context. For a large transport infrastructure investment the impacts to be covered will consist of: construction and maintenance costs, time savings, operation costs, accident savings, noise emissions, local air pollution and climate effects based on change in CO_2 emissions. These effects can with the exception of construction costs and time savings be determined in a relatively precise manner on the basis of current transport engineering modelling knowledge represented first and foremost by traffic and impact models (Leleur, 2000) (Danish Ministry of Transport, 2003).

Recently a methodology to handle construction costs and time savings that relate to demand prognoses has been set out with the Reference Scenario Forecasting (RSF) technique presented in (Salling & Leleur, 2009) (Salling & Leleur, forthcoming). RSF concerns Monte Carlo simulation and is based on prospect theory developed by Kahneman & Tversky in 1979 (Daniel Kahneman received the Nobel Prize in Economics in 2002 for his work in collaboration with Amos Tversky (1937-1996)). The prospect theory gave rise to Reference Class Forecasting (RCF), which was used by Flyvbjerg and others to set out various "uplift" principles (Priemus et al., 2008). This again laid the foundation for a scenario-based simulation procedure that made it possible to formulate and introduce Reference Scenario Forecasting (RSF) (Salling & Leleur, 2009). This paper about SIMDEC concentrates on applying RSF to provide input to MCA as this methodology (based on scenarios related to the actual study) for each of the examined alternatives can produce what has been termed a certainty graph (CG). This graph is made up of the probability estimates of achieving at least the benefit-cost rate (BCR) indicated as argument; thus $CG(x) = Prob(BCR \geq x)$. Examples are given in the following section about case illustration.

Multi-criteria analysis

The MCA is carried out by using the REMBRANDT technique (Olson et al., 1995) (Barfod et al., 2009) based on pair-wise comparisons for rating of the alternatives and determination of the criteria weights. The REMBRANDT technique is recognised as a both valid and practical framework (Lootsma, 1999) (Belton & Stewart, 2002) for which reason it was chosen to be used in this context. In SIMDEC a set of relevant decision criteria for the decision problem at hand is laid down. Generally such a set will consist of both monetary and non-monetary criteria. As already indicated the monetary criteria are taken into account by using a cost-benefit analysis, which again is used as an input to conducting a feasibility risk assessment (FRA) of each of the alternatives. Specifically a certainty graph, CG(x), is produced for each of the alternatives. For each alternative this graph represents its FRA-performance (based on the monetary criteria and the conducted Monte Carlo simulation). This FRA-performance is added as a criterion to the formulated non-economic criteria that typically represent strategic issues and impacts relating to the decision problem. Thereby the total criteria set is established for the examination of the decision problem. This set should be scrutinised to reduce possible overlapping with regard to criteria definitions, while at the same time it should be ensured that no valuable information for the decision-making has been left out.

The multi-criteria analysis proceeds by making pair-wise comparisons (either by the decision-makers themselves or facilitated by analysts interpreting information revealed by the decision-makers about their preferences). In the rating alternatives are successively compared two by two demanding a preference statement of the following type: very strong preference for.., strong preference for.., definite preference for.., weak preference for.., and indifference.. (Olson et al., 1995). A numerical REMBRANDT scale value associated with each statement is fed into the model and afterwards the same procedure is conducted for the other criteria to determine all the ratings and for the set of criteria to determine weights. Based on ratings and criteria weights REMBRANDT finally produces a total score (linear, additive summing-up based on the rates and weights) which makes it possible to rank the alternatives in accordance with their attractiveness.

One of the new features of SIMDEC is the mixing of monetary and non-monetary decision criteria with the first type of concern represented by the calculated FRA-performance. In SIMDEC the FRA-performance ratings of the alternatives are based on a set of pair-wise comparisons of the previously determined certainty graphs with each of these representing one of the alternatives that candidate for a decision about being selected and implemented. The final information presented to the decision-makers to base their decision on consists of the overall ranking based on the wider criteria set, where also the non-monetary aspects have been rated based on pair-wise comparisons. Below it is illustrated how SIMDEC manages to reduce very complex decision-related information to a set of criteria with rated alternatives that are used to produce a final ranking of the examined alternatives and thereby to indicate which alternative ought to be preferred.

CASE ILLUSTRATION

The case examines four alternatives for a new fixed link between Helsingør (Elsinore) and Helsingborg, see Figure 1, using SIMDEC (Larsen & Skougaard, forthcoming).

Figure 1: HH-Link Location at Helsingør (Elsinore)-Helsingborg (from Google Maps)

The four alternatives are listed in Table 1 with indication of type of construction and total cost (1 US$ equals around 5 DKK).

Table 1: The Four Alternatives incl. Cost in bn DKK for the HH-Fixed Link

HH-fixed link (alternatives)	Description (type of construction)	Total cost (bn DKK)
Alternative A1	Tunnel for rail (2 tracks) passenger traffic only	7.7
Alternative A2	Tunnel for rail (1 track) goods traffic only	5.5
Alternative A3	Bridge for road and rail (2x2 lanes & 2 tracks)	11.5
Alternative A4	Bridge for road (2x2 lanes)	6.0

Due to the high influence on the further development of the Øresund region a wider set of decision criteria has been adopted to lay the foundation for a comprehensive assessment of the four alternatives, see below:

Criterion 1: Robustness of feasibility (FRA-performance)
Criterion 2: Impact on towns
Criterion 3: Impact on ecology
Criterion 4: Impact on regional economics
Criterion 5: Impact on transport network and accessibility

Based on the Danish national manual for socio-economic assessment (Ministry of Transport, 2003) the benefit-cost rate (BCR) values shown in Table 2 have been determined applying transport modelling for road and rail traffic set in a context of three economic development scenarios spanning high, middle (continuation of established trend) and low economic growth (Larsen & Skougaard, forthcoming).

Table 2: The Four Alternatives with Conventional Cost-Benefit Rates

Cost-benefit rates for the four alternatives related to scenarios			
The four alternatives for the link	Economic growth expressed by three scenarios		
	HIGH	MIDDLE	LOW
A1: Tunnel, rail passenger service	1.71	1.51	1.32
A2: Tunnel, rail goods transport	0.24	0.18	0.12
A3: Bridge, road & rail	3.07	2.72	2.41
A4: Bridge, road	3.49	3.09	2.74

After this conventional socio-economic calculation Reference Scenario Forecasting (RSF) is applied to produce three sets of certainty graphs consisting of CG(A1), CG(A2), CG(A3) and CG(A4), with one set for each of the three scenarios mentioned above, see Figure 2 for the four CGs in the middle scenario.

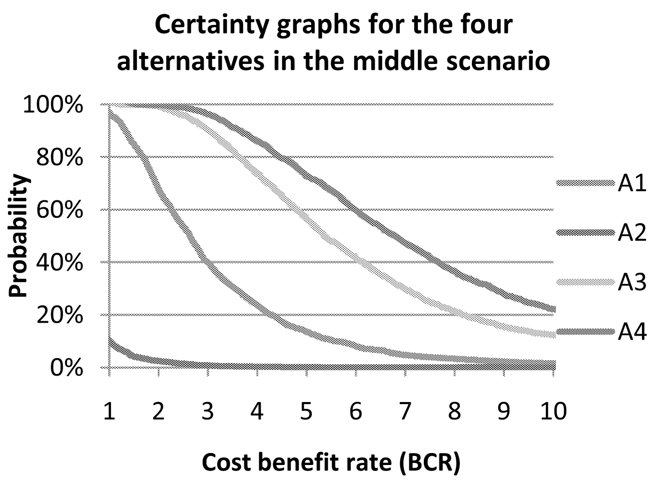


Figure 2: Certainty Graphs for the Four Alternatives in the Middle Scenario

The RSF-calculations behind the CGs are based on cost-benefit analysis and Monte Carlo simulation using estimated RSF distributions (Erlang and Beta-Pert distributions for construction costs and time savings respectively with the latter influenced by the actual scenario), see (Salling & Leleur, 2009). Each CG can be interpreted as follows: For x = 1 the probability or certainty of BCR ≥ 1 (with 1 indicating the ordinary socio-economic cut-off value with regard to feasibility) can be seen from the y-axis value. For the alternatives in the middle scenario in Figure 2 their certainty values (CVs) are: CV(A2) = 10%, CV(A1) = 97%. For the alternatives with CV ≥ 100% the certainty "distance" above 100% is added, so a BCR ≥ 2 with a probability of 100% (a value 100% higher than the ordinary cut-off value BCR = 1) is indicated as 100% + 100% = 200%. In this way the last two values in the middle scenario are obtained: CV(A3) = 100% + 23% = 123% and CV(A4) = 100% + 57% = 157%. In the scenario runs the expected economic growth associated with the actual scenario will affect the obtained certainty values, see Table 3

Table 3: The Four Alternatives also expressed by Certainty Values

Cost-benefit rates and certainty values for the four alternatives related to scenarios			
The four alternatives for the link	Economic growth expressed by three scenarios		
	HIGH	MIDDLE	LOW
A1: Tunnel, rail passenger service	1.71; 98 %	1.51; 97 %	1.32; 94 %
A2: Tunnel, rail goods transport	0.24; 14 %	0.18; 10 %	0.12; 7%
A3: Bridge, road & rail	3.07; 149 %	2.72; 123 %	2.41; 116 %
A4: Bridge, road	3.49; 206 %	3.09; 157 %	2.74; 142 %

Where the conventional cost-benefit rate gives a deterministic point estimate of the feasibility, the RSF-based certainty values give a probability based interval estimate of how the two most important uncertainty factors could affect such a point estimate. Specifically, construction costs and time savings are simulated using historical reference class knowledge made operational by using the Erlang and Beta-Pert distributions respectively with the latter embedded in a scenario context (Ibid.). In the simulation the uncertainty due to the estimation of construction costs is considered generally, i.e. across the scenarios and not related to a specific scenario.

In SIMDEC the certainty graphs and certainty values are used as the basis of the final REMBRANDT procedure with regard to criterion 1 about robustness of feasibility. The four alternatives are compared two by two resulting in altogether (4 x 3)/2 = 6 pair-wise comparisons as shown in Table 4.

For the remaining four criteria information has been gathered to serve as sufficient background for the criteria rating, which leads to additional 4 x 6 comparisons. As an example the pair-wise comparisons for criterion 5 about

Table 4: REMBRANDT Rating of Criterion 1: Robustness of Feasibility

Pair-wise comparison of alternatives under criterion 1: Robustness of feasibility				
Scale value (j,k)	A1	A2	A3	A4
A1	0	Strong (+6)	Definite (-4)	Strong (-6)
A2	Strong (-6)	0	Strong (-6)	Very strong (-8)
A3	Definite (+4)	Strong (+6)	0	Weak (-2)
A4	Strong (+6)	Very strong (+8)	Weak (2)	0

Note relating to j compared to k: Indifference 0, weak +2, definite +4, strong +6 and very strong +8. Observe that a reversal of j to k is indicated by –. Elements in the diagonal are all necessarily 0.

impact on transport network and accessibility are indicated below, see Table 5.

Table 5: REMBRANDT Rating of Criterion 5: Impact on Transport Network and Accessibility

Pair-wise comparison of alternatives under criterion 5: Network and accessibility				
Scale value (j,k)	A1	A2	A3	A4
A1	0	Strong (+6)	Definite (-4)	Weak (-2)
A2	Strong (-6)	0	Very strong (-8)	Strong (-6)
A3	Definite (+4)	Very strong (+8)	0	Definite (4)
A4	Weak (+2)	Strong (+6)	Definite (-4)	0

The rating values for all five criteria are shown in Table 6 together with the weights determined by pair-wise comparison of the five criteria with the latter demanding (5 x 4)/2 = 10 comparisons. Thus altogether, as the rating of each of the five criteria demands 6 comparisons, the values in Table 6 are determined by (5 x 6) + 10 = 40 pair-wise comparisons carried out by using the REMBRANDT scale and fed into the SIMDEC model. As concerns the theoretical set-up of REMBRANDT it should be noted that the processing of scale-values for the determination of ratings and criteria weights differs with regard to the so-called progression factors, see (Olson et al., 1995). This, however, does not influence the easy and straightforward application of this multi-criteria methodology.

It should be observed that a documentation report, the SIMDEC 'logbook', is worked out as part of doing the pair-wise comparisons. Afterwards the considerations behind each comparison can be studied/inspected to judge the overall validity of the model outcome. In case of disagreement and debate among the decision-makers this

Table 6: REMBRANDT Rates and Weights for the Five Criteria

REMBRANDT rates and weights determined by using altogether 40 pair-wise comparisons					
Five criteria / four alternatives	A1 rates	A2 rates	A3 rates	A4 rates	Weights
Robustness of feasibility	0.03	0.00	0.19	0.78	0.35
Impact on towns	0.20	0.79	0.00	0.01	0.13
Impact on ecology	0.47	0.47	0.03	0.03	0.06
Impact on regional economics	0.05	0.00	0.76	0.19	0.19
Impact on network and accessibility	0.05	0.00	0.84	0.11	0.27

also makes it possible to make adjustments as a basis for model re-runs.

Based on the values in Table 6 the total score for each alternative is determined using all five criteria by Σ(weight x rate) leading to the prioritising of the four alternatives shown in Table 7.

Table 7: The Four Alternatives with Total Scores and Rank Order indicated

HH-fixed link (alterna-tives)	Description (type of construction)	Cost (bn DKK)	Score	Rank
A1	Tunnel for rail (2 tracks) passenger traffic only	7.7	0.09	4
A2	Tunnel for rail (1 track) goods traffic only	5.5	0.13	3
A3	Bridge for road and rail (2x2 lanes & 2 tracks)	11.5	0.44	1
A4	Bridge for road (2x2 lanes)	6.0	0.34	2

DISCUSSION

Conventional decision support for deciding upon the four HH-alternatives would consist of a calculation of benefit-cost rates (BCRs) supplemented by various information not accounted for in the BCRs. SIMDEC offers an approach where simulation and multi-criteria analysis are applied to deal with the complex decision problem. The BCR information contained in Table 2 indicates that three (A1, A3 and A4) out of the four alternatives are socio-economically sound with BCR values also in the low growth scenario being well above 1. Conducting afterwards reference scenario forecasting (RSF) however indicates, by inspecting the produced certainty graphs (CGs) and certainty values (CVs), that really only two alternatives (A3 and A4) are sound when including risk analysis in the assessment. By accounting for estimation bias in the cost estimates and prognosis bias in the forecasting of traffic and exploring this by simulation embedded in scenarios the alternative A1 is shown to have a feasibility that cannot be considered robust.

With alternatives A3 and A4 remaining as candidates for implementation these are explored in a wider context, where strategic, non-monetary issues are introduced together with the criterion about robustness of feasibility based on the described risk analysis. With CGs and CVs of the alternatives as input to this criterion a REMBRANDT multi-criteria analysis is carried out which comprises also the criteria about the impacts on towns, on ecology, on regional economics and on transport network and accessibility. The result is that even with the highest criterion weight on robustness of feasibility the order of importance with regard to A4 and A3 is now reversed as alternative 3 appears as the most attractive alternative.

For the lower ranking alternatives A1 and A2 it should be observed that A2 is now better than A1. The wider assessment based on the multi-criteria analysis has thus revealed and indicated some qualities contained in A2, which were not captured by the BCR values.

CONCLUSION AND PERSPECTIVE

The SIMDEC approach is seen as promising since relatively complex decision problems of a strategic nature can be based on both explicit risk precaution and influence from a set of wider, non-monetary issues. As the approach with its successive building up of assessment information is easy to grasp it can be applied in decision sessions where a high decision-maker involvement is possible. SIMDEC software has been worked out that can facilitate such decision conferences. Hereby the SIMDEC methodology becomes embedded in a process that includes also criteria formulation by use of soft operations research methods such as brainstorming and futures workshops; due to the software the participants can ask various 'what-if'-questions to test the robustness of the priority-ordering of the alternatives. One important issue that can be treated is how different stakeholder strategies – defined by the chosen set of decision criteria and the stated value inputs to the pair-wise comparisons – will affect the result (Leleur, 2008) (Jeppesen, forthcoming).

The SIMDEC approach has so far been tested on transport planning problems but the perspective is to explore its potential also for complex decision problems outside the transport sector. It is expected that alternatives for construction projects in general can be examined by SIMDEC in a way that satisfies both theoretical validity and practical user-friendliness.

REFERENCES

Banister, D. & Berechman J. (2000). *Transport Investment and Economic Development*. UCL Press.

Barfod, M.B., Jensen, A.V. & Leleur, S. (2009). "Examination of decision support systems for composite CBA & MCA assessments of transport projects". Paper presented at the 20th International Conference on Multiple Criteria Decision Making, Chengdu, China, June 2009.

Belton, V. & Stewart, T.J. (2002). *Multi Criteria Decision Analysis – An Integrated Approach*. Kluwer Academic.

Danish Ministry of Transport, 2003. *Manual for socio-economic analysis – Applied methodology in the transport sector*. Copenhagen (in Danish).

Jeppesen, S.L. (forthcoming). "Sustainable Transport Planning". Ph.D. thesis, Department of Transport, Technical University of Denmark.

Larsen, L.A. & Skougaard, B.Z. (forthcoming). "Vurdering af alternativer for en fast forbindelse Helsingør-Helsingborg". M.Sc. thesis, Department of Transport, Technical University of Denmark (in Danish).

Leleur, S. (2000). *Road infrastructure planning – A decision-oriented approach*. Second Edition, Polyteknisk, Copenhagen.

Leleur, S. (2008). *Systemic planning – Principles and methodology for planning in a complex world.* Second Edition, Polyteknisk, Copenhagen.

Lootsma, F.A. (1999). *Multi-Criteria Decision Analysis via Ratio and Difference Judgement*. Kluwer Academic.

Olson, D.L., Fliedner, G. & Currie, K. (1995). "Comparison of the REMBRANDT system with analytic hierarchy process". *European Journal of Operational Research,* Vol. 82, Elsevier., 522-539.

Priemus, H., Flyvbjerg, B. and Wee, B. van, (2008). *Decision-making on Mega-Projects: Cost-Benefit Analysis, Planning and Innovation.* Edward Elgar Publishing.

Salling, K.B. (2008). "Assessment of Transport Projects: Risk Analysis and Decision Support". Ph.D. thesis, Department of Transport, Technical University of Denmark.

Salling, K.B. & Leleur, S. (2009). "Modelling of Transport Project Uncertainties: Risk Assessment and Scenario Analysis". In proceedings of the Modelling and Applied Simulation '09 conference, Puerto de la Cruz, Tenerife, September 2009.

Salling, K.B. & Leleur, S. (forthcoming). "Reference Scenario Forecasting: A New Approach to Transport Project Assessment". Paper accepted for presentation at the 12th World Conference on Transport Research in Lisbon, July 2010.

Vose, D., 2002. *Risk Analysis – A quantitative guide*. Second Edition. John Wiley and Sons.

BIOGRAPHY

Steen Leleur is professor of decision support systems and planning and head of the Decision Modelling Group (DMG) at the Technical University of Denmark, Department of Transport. Over the years Steen Leleur has been involved in many national and international research projects, among others several within the European Commission's strategic transport research programmes. Currently he is the leader of subtask 5 about modelling in the research project "Uncertainties in Transport Project Evaluation" (UNITE) 2008-2012; UNITE is funded by the Danish Research Council.

Simulation of a Health Care Knowledge-based System with RFID-generated Information

Yannick Meiller[1], Sylvain Bureau[1], Wei Zhou[1], Selwyn Piramuthu[2]
[1] Information Technologies & Modeling, ESCP Europe, Paris, France
[2] Information Systems and Operations Management, University of Florida, USA
email: yannick.meiller@novapten.com, {sylvain.bureau, wzhou}@escpeurope.eu, selwyn@ufl.edu

KEYWORDS
Simulation, health care, RFID, knowledge-based system

ABSTRACT

Health care organizations are under increased pressure to continually increase their operational efficiency while decreasing the overall operating costs with no appreciable decrease in the delivered quality of health care. Given the nature of such an environment where lives are at stake, it is natural to operate under a larger safety factor where risks are kept close to their necessary minimum. RFID tags are increasingly being used in health care organizations to reduce errors and to generally increase the effectiveness of the processes. We consider management of bottled gas delivery within a health care environment through a knowledge-based decision support system using item-level information generated through RFID tags. We use simulation to study the underlying dynamics.

Introduction and Background

Health care organizations, like most other organizations, operate under resource constraints that seem to get tighter every year. However, unlike some organizations that operate in very low risk environments, health care environment is characterized by its close proximity to invaluable human lives where stakes are extremely high. This translates to the constant need to maintain a high level of operational effectiveness and quality in health care delivery even when efficiency, cost, schedules, etc. are overrun (Tu, et al., 2009). While this works well for a short time-span, it is difficult to sustain such an operating environment without improvements in efficiency in the work environment. Over the past several years, there have been many successful attempts to approach this and related issues where the introduction of automation has reduced routine errors and increased the operating efficiency. We consider a facet of operations in a hospital environment, specifically the one that is related to the delivery and management of bottled gas. While RFID tags are not yet commonly being used in such applications, we incorporate RFID tags on bottled gas and study their dynamics in a hospital environment with respect to their efficient utilization.

The rationale for the choice of this (i.e., bottled gas in a hospital setting) application is two-fold: RFID tags are easily applied to these bottles, and there is an urgent need to track and trace these bottles. We study the management of bottled gas at a French hospital where these bottles enter the hospital environment primarily through the hospital's pharmacy. We ignore other means of entry (e.g., arrival of patients with bottles in ambulance or transferred from other hospitals) although these could add up to a significant number over time. The hospital pharmacy then distributes these bottles as per the dictates of internal (to the hospital) demand.

The primary impetus for this study is the current lack of traceability of bottles at this hospital (Meiller and Bureau, 2009). For example, surgery departments routinely over-order these bottles, which are then buffered locally and used as necessary. The left-over bottles (i.e., those that were never opened to be used) are kept within the department for the next time, and the used bottles are rarely returned on time to the pharmacy. I.e., the used bottles remain at the surgery department much longer than necessary. While this may not be a problem from a medical perspective, it is an issue since the pharmacy has to pay rent on these bottles for every day they are in the hospital. While paying rent when a bottle is full is necessary, it is a waste to pay rent on an empty bottle. Affixing RFID tags on these bottles would help inform the pharmacy when they are used and when it's time for them to be returned to the pharmacy.

In addition to the internal misplacement of used bottles, two other scenarios lead to the misplacement/loss of bottles from the pharmacy. One of them occurs when a patient gets transferred to another hospital and the bottle currently being used by this patient is transported along with the patient to the other hospital. These bottles are rarely returned to the original hospital. The other occurs when an ambulance personnel removes a bottle from the hospital premises without the permission of the pharmacy personnel. These bottles, again, are rarely returned when empty.

The remainder of the paper is organized as follows: we present the proposed knowledge-based framework for addressing missing/misplaced gas bottles in the next Section. We then illustrate the proposed framework us-

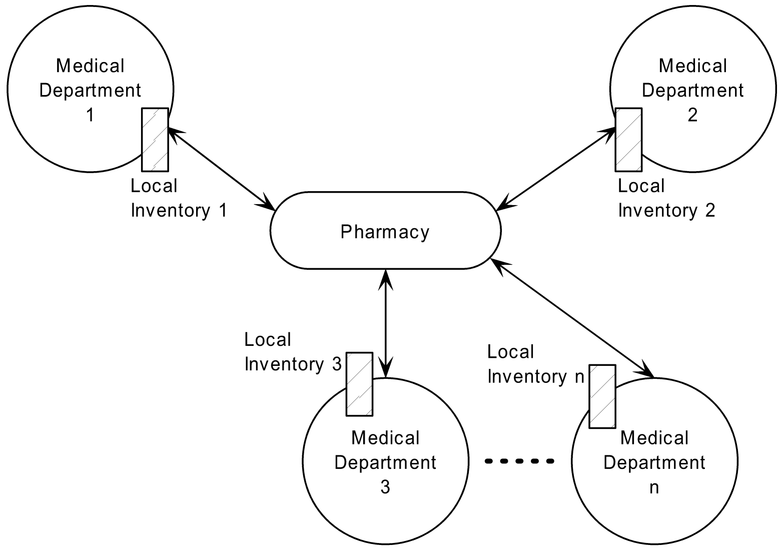


Figure 1: HealthCare Inventory Management: An Overview

ing simulation in the following Section. We later conclude with a brief discussion.

The Proposed Knowledge-based Framework

Figure 1 illustrates the structure of inventory management at a French hospital. The pharmacy performs the primary role as the source of medical equipment supply to the different medical departments within the hospital, and deals with entities such as filled oxygen bottles, surgical equipment, etc. When there is demand for medical equipment at one of the local medical departments, a staff member checks for the availability of local inventory. If the equipment needed is not available in the department, the staff obtains the required items from the pharmacy. If the item is not available at the pharmacy but is available at another department within the hospital, an urgent request is made and a pharmacy staff member is sent to retrieve the item from local inventory in another medical department.

While much of the process is done manually, there is tendency for the staff at each of the departments to hoard more than what they currently need due to several possible reasons including (a) fear of out-of-stock at the pharmacy-level, (b) preference for holding and managing stocks locally as a buffer at the department level, or (c) general laziness to order from the pharmacy every time need occurs. Moreover, in addition to stocks unnecessarily remaining idle at the department levels, an appreciable number of empty bottles are not returned to the pharmacy as soon as they become empty. This leads to unnecessary payment of rent on these bottles.

We propose an automated framework, which is enabled by real-time RFID item-level information (e.g., Zhou, 2009) on medical equipment, to facilitate the operation and to improve the effectiveness of health care inventory management and improve the overall medical service quality while reducing associated costs. Because human health and possibly even life is at stake in such a health care setting, time spent on medical preparation play an important role on the quality of medical service provision. The cost of medical inventory control includes both the inventory cost and labor cost. Moreover, in public health care institutions both physical resources and medical labor resources are highly constrained, there is a need to balance resource allocation from all perspectives.

The proposed framework in consort with tagging all gas bottles is aimed at generally reducing inventory-related costs primarily through effective management and accountability. This is accomplished by automating a majority of the process and providing complete visibility of these items (both empty and full gas bottles) at all times including when they are being used as well as those that remain idle in inventory. From availability of complete visibility of each of the bottles, it is relatively easy to take inventory from pharmacy's perspective since these bottles are either in the hospital in which case its location is known or outside the hospital and its location (i.e., who took it, when, to where, and the contact information of its current location so it can be retrieved upon its use). Based on data obtained from continually polling the tagged bottles, the pharmacy can also determine the statistics on how often each of the bottles sit around once it leaves the pharmacy and before being eventually used. Moreover, statistics on the amount of time taken by each of the bottles from time they are empty to the time they are back at the pharmacy is very useful for the Pharmacy in reducing unnecessary rent payments on empty bottles.

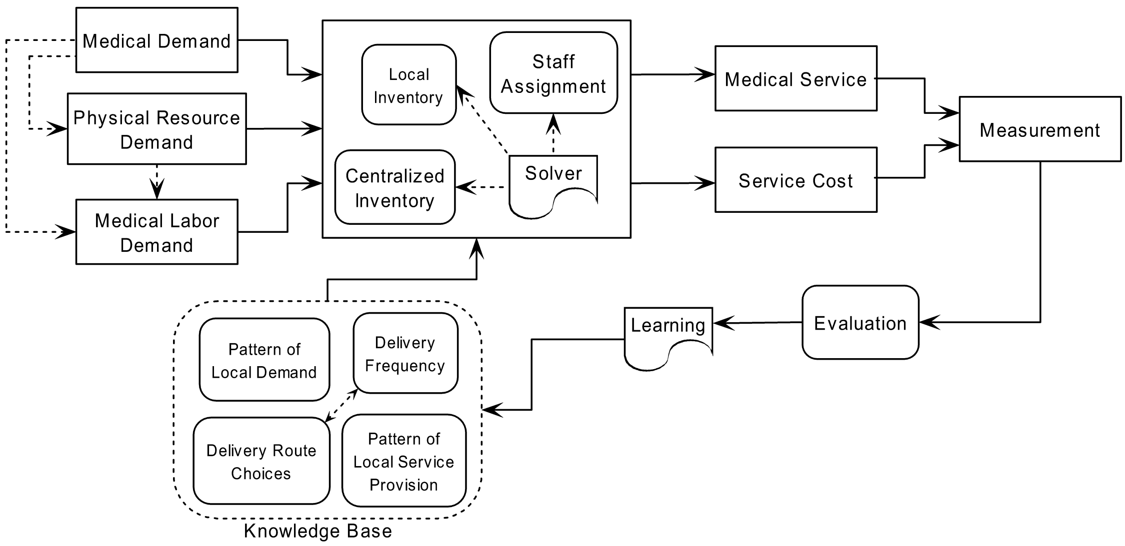


Figure 2: RFID enabled Adaptive Learning Framework for HealthCare Inventory Management

The proposed framework is presented in Figure 2. The exogenous factors include the actual medical demands that further become the demand on both physical and on medical labor resources. With these inputs, the problem solver determines the appropriate local and centralized inventory, and effective assignment of staff in charge of inventory management. After a medical service is performed, both service quality and associated cost are determined and evaluated. Based on evaluation, associated benchmarks, and the extent of possible potential improvements, appropriate remedial measures are recommended and implemented.

Moreover, the proposed framework includes the ability to learn from past experience. Here, effective inventory management is rewarded while poor management is not. This reinforcement of good behavior as well as adoption of better means to approach inventory management results in continual improvement of the system performance over time. Thanks to RFID-generated information on all tagged items (here, gas bottles), it is now possible to map the presence of these bottles throughout the hospital environment. It is also possible to identify those that are new and those that are used. Once opened, a content of the bottle is used as much as is necessary for that given application and the remaining gas is returned in the bottle to the pharmacy. General estimates are that on an average the bottles are returned with about 40% of the gas still remaining. This 40% is considered waste and is never re-used. We do not consider this wastage in this study, although means to salvage this wastage could only improve the effectiveness of gas use in the hospital system.

For the cases where the gas bottles are taken outside of the hospital premises, such as when a patient is transferred to another hospital or when an ambulance leaves the hospital premises with a few of these bottles, the proposed system keeps track of when, where, and by whom a given bottle was taken and arrangements are made off-line to promptly retrieve these bottles as they are no longer required by the borrower. While this can also be done on-line using the proposed system (e.g., send frequent electronic reminders asking the borrower to return the borrowed bottles), we do not explicitly consider this facet. Our model implicitly assumes that the bottles somehow make their way back to the lending pharmacy on time without getting lost in transit.

While it is relatively easy to accomplish this task in a static environment, it is not straight-forward in a dynamic setting that is rife with uncertainties. The proposed system is specifically designed to handle uncertainties and perform well in a dynamic environment where there is a need to react both proactively as well as reactively as the situation dictates. The proposed system accomplishes this reflex by observing the system dynamics both from internal and external perspectives and learns to make the best decision for any pattern exhibited by an instantaneous snapshot of the system. This ability to observe an instant snapshot of the system, again, is facilitated by the item-level information that is generated by the RFID tags. Not only are the RFID tags able to generate appropriate information on their substrate objects, they are also able to transfer this information instantaneously in a continual manner without any interruptions. Without RFID tags and their small footprint and cost, this may not be possible.

The Measurement, Evaluation, and Learning components work in consort to synergistically determine any deficits in knowledge stored in the knowledge base. The knowledge base is periodically updated when deficits are present. Generally, these deficits are determined based

# of Entries/Arrival	# Bottles	Cost[60/100]	Cost[120/200]	Cost[200/500]
Lognormal(2,0.4)	529	1.9321	1.8584	1.7875
Lognormal(3,1)	893	2.1357	2.0551	1.9789
Lognormal(5,1)	1633	2.1518	2.0722	1.9954
Lognormal(7,1)	2385	2.1272	2.0461	1.9696
Lognormal(9,1)	3117	2.2574	2.1788	2.1034
Lognormal(10,2)	3477	2.1896	2.1088	2.0320

Table 1: Results for cases without RFID

# of Entries/Arrival	# Bottles	Cost[60/100]	Cost[120/200]	Cost[200/500]
Lognormal(2,0.4)	529	0.1767	0.1030	0.0321
Lognormal(3,1)	893	0.1909	0.1103	0.0341
Lognormal(5,1)	1633	0.1904	0.1108	0.0339
Lognormal(7,1)	2385	0.1922	0.1111	0.0345
Lognormal(9,1)	3117	0.1878	0.1092	0.0338
Lognormal(10,2)	3477	0.1921	0.1113	0.0345

Table 2: Results for cases with RFID

on a snapshot of its instantaneous performance at any given point in time. Appropriate examples addressing these deficits are identified and the knowledge base is incrementally updated to reflect identified deficits (e.g., Park, et al., 2001). Each of these components play important roles in this process since without measuring and evaluating current performance, it is impossible to know what is missing and what is yet to be learned and what can be learned. The Learning component is the primary source of input to the knowledge base, and it directly updates learned knowledge in the knowledge base whenever it is deemed necessary.

The knowledge base is the core of the proposed framework, and is used to assist with future medical inventory management. The knowledge base includes patterns of medical demand in each department, patterns of service provision, logistical route choices and feasible delivery frequency. The choice of logistical routes and the delivery strategy are mutually dependent.

Simulation Results

To illustrate the beneficial aspects of RFID tags for tracking and tracing gas bottles in a hospital context, we simulated the essence of this environment with and without the incorporation of RFID tags in the system. While we do not consider all possible benefits of using RFID tags in such a system, we are primarily interested in the cost (or, rather, cost savings) aspect of these systems when RFID tags are introduced for tracking and tracing purposes. From our past experience with simulating real-time systems, we simulated the system using several arrival rates and inventory reordering amounts to reduce bias due to chosen parameters.

We simulated the system for 365 days after a warm-up period of seven days, and the results are provided in Tables 1&2. The number of entries per arrival (or, per time period) is generated using the general distribution as given in the column titled "# of Entries/Arrival." Here, the number of entries represents the number of requests for bottles or rather, more accurately, the demand for bottles during each time period from both within the hospital as well as from outside (e.g., ambulance, transferred patients, etc.). This may not necessarily translate into request for bottles at the pharmacy level since some of these demands could be met by local (i.e., the department level) buffer stock. Upon use, bottles are returned to the pharmacy in a staggered fashion by the different departments, as in the real-world hospital environment.

The number of bottles that passed through the system during the 365 days is given in the column titled, "# Bottles." Clearly, this column depends on the type of distribution used in generating bottle arrivals as given in the first column. The next three columns represent data for individual time units (i.e., days) and not the entire 365 days as in the second column.

We modeled the cost of bottles in normalized form where each filled bottle costs one currency unit for gas and rent payment and each empty bottle costs 0.02 currency units of rent payment. The last three columns in Tables 1 & 2 refer to the cost per time unit associated with only the bottles (cost of gas + bottle rental cost), where 60/100 represents total inventory of 100 units and order placement (to top up inventory to 100) when inventory falls below 60 units as in an EOQ (Economic Order Quantity) model.

Ideally, the requestor of gas bottles should be able to request bottles only when they are needed and once they have made a request, the item(s) would ideally be delivered instantaneously. Although this idealistic Just in Time (JIT) scenario does not exist in reality, the proposed system allows for keeping the local inventory (buffer) at the departmental levels to a minimum while allowing for the possibility of ordering exactly what is needed and getting the order fulfilled on time from the pharmacy. While this ideal from the perspective of the ultimate customer (e.g., surgeon) of these gas bottles, it is not feasible at the pharmacy level since the pharmacy generally incurs an ordering cost for each order placed. Moreover, it is difficult to predict the demand at the pharmacy level relative to those at individual department levels since the former is a compilation of the entire set of the latter.

Given the sensitive nature of this domain, the pharmacy order-placement generally tends to be on the conservative end where more than enough safety stock is always maintained to be ready for any emergency situations. To discourage frequent order placement and to reduce order placement cost, we include a nominal cost of 0.5 currency unit for each placed order. This barely leaves a dent when the arrival frequency of requests for gas bottles is high since this only amounts to 0.5 currency units per time period (e.g., day) when an order could possibly be placed. When the number of bottles requested per year is 529 (as in the lognormal(2,0.4) case), when order for bottles is placed by the pharmacy every day, it amounts to a reasonable sum of 182.5 currency units which is about a third of the filled bottle cost for the year. On the other extreme, among the arrival rates considered, when the number of bottles requested per year is 3477 (as in the lognormal(10,2) case), even when an order is placed by the pharmacy every day (182.5/3477 = 0.05), it only amounts to about 5 percent of the total amount. While an order cost of 5 percent is still not a trivial amount, it can be reduced even further through appropriate modification of order frequencies.

As can be seen from Tables 1&2, the cost values decrease as the number of items in inventory increases. This is primarily due to the average reduction in ordering cost per item which is a direct result of reduced ordering frequency. However, we expect this to be reversed with an increase in the unit cost of gas bottle with respect to order placement cost. Since determining the optimal order quantity or cost structure is not the purpose of this study, we do not concern ourselves with this part of results. The results also include information on the primary purpose of this study, and show the relative savings when RFID tags are used. Although the results are biased based on the parameter values we used, the general results hold regardless of the values used simply because of the accuracy with which part inventory-taking can be accomplished thanks to the overall visibility and availability of item-level information and also the reduction in wastage (e.g., unnecessary rental payment on empty gas bottles, unnecessary inventory because the exact current stock level is unknown without RFID-generated information).

There are, clearly, other benefits to RFID-tagged bottles in addition to the obvious ones we discussed such as the ability to precisely locate and in turn track and trace the tagged items at any point in time, accurate inventory management, reduced wastage, etc. Knowing where any given bottle is could possibly help the medical staff access it without much searching, the bottles can be used in the order in which they were delivered to avoid any problems due to staleness, etc., and it is relatively easy to locate and identify specific bottles when any batch of such bottles are recalled due to manufacturing or other defects, among others. While it is relatively straight-forward to identify a recalled item from a set of items that are readily available, accomplishing the same when items are misplaced is rather challenging since the instance-level information on a misplaced bottle is generally not easily retrieveable, which is not the case when these bottles are rfid-tagged.

Discussion

We set out to study the dynamics of gas bottle inventory management in a health care setting and the effect of RFID tagging these bottles. We proposed a knowledge-based framework and considered the example of a French hospital to illustrate our framework. Preliminary results show that RFID tags beneficial in several respects: (1) they result in an overall reduction in inventory cost over time, while (2) providing complete visibility on the existence of every tagged item (here, gas bottles - both empty and full) continuously over time. While (1) is certainly good, (2) provides useful information that the pharmacy inventory manager could utilize to effectively manage gas bottle inventory in the hospital. Combine this with the ease of dealing with recalled items, effective use of items before their expiry date, etc., it is not difficult to appreciate the effectiveness of RFID tagging bottles in hospital environments.

REFERENCES

[1] Meiller, Y. and S. Bureau. 2009. Logistics Projects: How to Assess the Right System? The Case of RFID Solutions in Healthcare. *Proceedings of the Americas Conference on Information Systems (AMCIS).*

[2] Park, S.C., S. Piramuthu, and M.J. Shaw. 2001. Dynamic Rule Refinement in Knowledge-based Data Mining Systems. *Decision Support Systems*, 31, 205-222.

[3] Tu, Y.-J., W. Zhou, and S. Piramuthu. 2009. Identifying RFID-embedded Objects in Pervasive Healthcare Applications. *Decision Support Systems*, 46, 586-593.

[4] Zhou, W. 2009. RFID and Item-Level Visibility. *European Journal of Operational Research*, 198(1) (October) 252-258.

MODELING URBAN SERVICES USAGE WITH SOCIAL INSECT ALGORITHMS

Rawan Ghnemat[(1)], Cyrille Bertelle[(2)], Gérard H.E. Duchamp[(3)]

[(1)]German-Jordanian University, PO Box 35247, Amman 11180, Jordan
email: rawan.ghnemat@gmail.com

[(2)]LITIS - Le Havre University, 25 rue Philippe Lebon, BP 540, 76058 Le Havre Cedex, France
email: cyrille.bertelle@gmail.com

[(3)]LIPN - Paris XIII University, 99 avenue Jean-Baptiste Clément, 93430 Villetaneuse, France
email: ghed@lipn.univ-paris13.fr

KEYWORDS

swarm intelligence, complex systems, self-organization, ant systems, urban dynamics, spatial organization

ABSTRACT

This work concerns the implementation of swarm intelligence models in order to study the spatial emergence of organizations within self-organized systems, under multi-criteria constraints. The target application concerns user/service urban systems. A social insect algorithm based on ant nest building metaphor is proposed to deal with the adaptive process involved in the model. Experiment outputs are given and a specific analysis methodology is developed in order to better understand the complexity of the studied phenomena.

PROBLEM SETTING AND OBJECTIVES

The impressive increase of economical, technological, social, and environmental changes in our world makes their management become a great challenge. Implementing wrong policies to solve such problems can make the problem become worse or can generate new ones. Many of the problems we have to face nowadays arise as non predictable side effects of our own past actions. Most of these wrong policies are not flexible or adaptive, and we cannot change them in order to achieve our goals within a dynamic solution environment. In other words, solving such problems by studying a part of them without modeling the complexity of the different parts can provide an efficient solution for short term but a negative one for long term, leading to non reversible system evolution.

Effective decision making and learning in a world of growing dynamic complexity requires us to search a new way of system modeling and to expand the boundaries of our mental models. We have to develop tools in order to understand how the structure of complex systems creates their own behavior.

Our purpose is to analyze organizations or societies within their spatial complexity. Solving this problem in sustainable way, should start by understanding how the individual behavior in the organization affects the whole system behavior. Observing complex self-organized systems in the nature (like social insects) and understanding them leads to discover concepts of emergence (1). The power of these systems does not come from any central control but from their flexible interactions with themselves and with their environment in an adaptive way. This problem involves complex networks of location interactions, complex networks of individual characteristics' interaction and even complex networks of multi-scale decision making, that are: the decision of individuals, the decision of services managers and the decision of society development planners.

The global objective of this work is to contribute to better understanding and exploring such complex systems using modelling, simulations and analysis. We develop in the following the objectives and the contributions of this work.

The aim of this study is to model and analyze social organizations from services-users interactions, respecting their complexity. Our objective is to highlight the emergent complexity based on the two main factors: (i) the spatial dimension as a major input for the component interaction, (ii) the self-organization as a major process for the system dynamics and adaptation.

In order to model the dynamics of complex interaction between users and services with various characteristics, we introduce two levels of description: the entities (users) level and the level of organizations (services), as described in figure 1. The objective is to implement an adaptive mechanism for our model structured threefold:

- *The emergent process.* The inputs are (i) the system of spatial users with various characteristics and preferences, (ii) the system of spatial services with various characteristics. The output is the spatial distribution of users over the services system. The process is based on an attraction mechanism involving complex interaction between users and services.
- *The adaptive process.* The input is the spatial distribution of users over the services system which is produced by the emergent process. The output is the response of the services according to their dynamic usage. This response leads to modify the characteristics of the services.
- *The feed-back process.* The input is the new characteristics of the services which are produced by the adaptive process. The output is the new spatial distribution of users according to the services evolution.

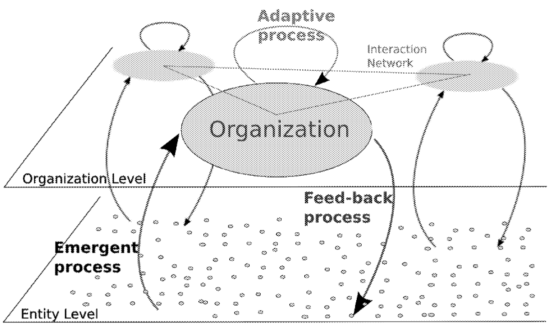


Figure 1: Users and Services spatial organizational model

This three faces model produces an efficient tool to study the complexity of the adaptation and evolution of the services-users systems (5; 4). The final objective is to produce relevant decision making tools based on such simulations and analysis.

In the following section, we describe the bio-inspired mechanism as the basis of our model. Then the model formalism will be described and finally experiments and analysis are given.

NATURAL TEMPLATE IN COLLECTIVE BUILDING

Termites mounds, ant nest, bees and wasp hives are some examples of a complex structures which emerge from the interactions between the social insects and their spatial environment. One of the mechanisms used in these structures is templates. The template is a pattern existing in the environment and used to construct another pattern (2). Temperature, humidity, chemical gradient, physical templates are different kinds of templates that we can find in different natural structures.

An example of template (combined with self organization mechanism) is the one proposed by Leptothomx albipennis ants which construct simple perimeter walls in a two-dimensional nest at a given distance from the tight cluster of ants and brood and which serves as a chemical or physical template (see figure 2). The probability of depositing a grain is:

- Higher when both the distance from the cluster is appropriate and the local density of grains is large.
- Lower when the cluster is either too close or too far and when the local density of grains is small.

When the distance from the cluster does not lie within the appropriate range, deposition can nevertheless be observed if grains are present. Conversely, if the distance from the cluster is appropriate, deposition can take place even if the number of grains is small.

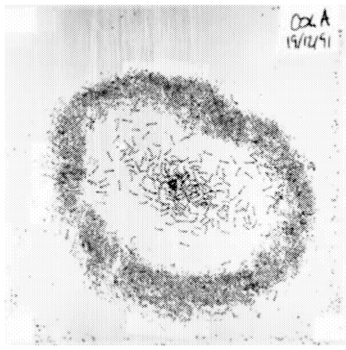


Figure 2: Wall formation generated by Leptothomx albipennis ant colonies (3)

Wall Building Model: Implementation in Repast

The algorithm is based on the ant clustering algorithm which consists to simulate how ants are able to spatially classify corpses or larvea with elementary decentralized process.

The algorithm is based on the following rules which are very simple behavioral rules that each ant implements :

- When an ant is moving without carrying yet material and finds some material, the ant will take the material respecting the probability number :

$$P_p = \left(\frac{k_1}{k_1 + f}\right)^2 \quad (1)$$

where f is the material density that the ant perceives locally around itself and k_1 is the threshold.

- When an ant is moving when carrying some material, the probability to deposit it is computed by :

$$P_d = \left(\frac{f}{k_2 + f}\right)^2 \quad (2)$$

where f is still the material density that the ant perceives locally around itself and k_2 is another threshold.

In order to extend this basic algorithm to nest building algorithm, we have to define a template probability p_t which will be added to the probability of picking up and dropping.

Using this template function, we replace in the clustering algorithm, the two previous probabilities of picking up and depositing a grain of material by new values integrating this template function.

The new probability of picking up a grain of material is:

$$P'_p = P_p(1 - P_t) \quad (3)$$

where P_p is the previous probability given by equation (1).

The new probability of depositing a grain of material is:

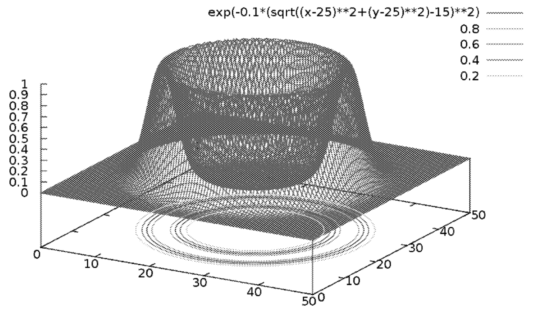
$$P'_d = P_d P_t \quad (4)$$

where P_d is the previous probability given by equation (2).

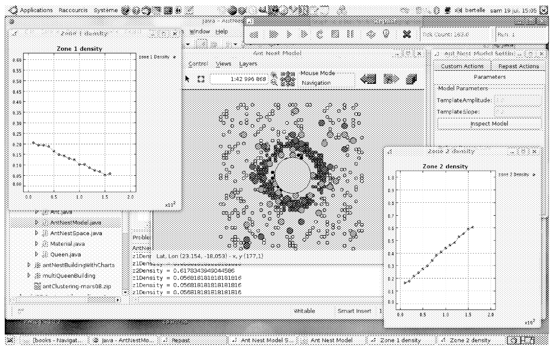
On figure 3, we represent a template function adapted to the natural ant wall building and we represent the result of a simulation made on the agent-based java platform, Repast (6). The material which has been deposited are here represented with blue color.

Adaptive Spatial Organization Feedback Implementation

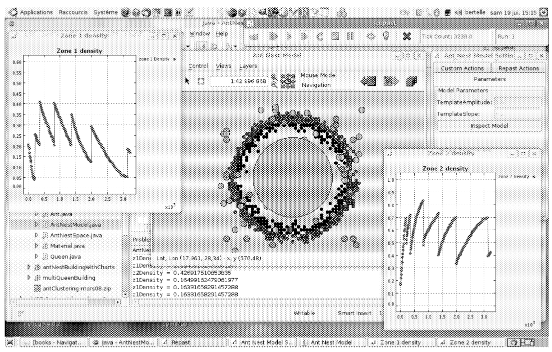
Complex systems deal not only with emergent organization processes from the interaction of its own entities, but also with the feedback processes of the organization over its own components. In the proposed model, we can take into account such feedback process and we present in figure 3 an adaptive process which makes the queen (describing the organization itself) modify the environment and the clustering processes itself. Following the template function, the queen locally defines around it two zones. The first zone is near itself and it is not expected to find material there. The second zone corresponds to the template maximum and it is expected to find a great concentration of material there. In the simulation, we count in a dynamical way the number of materials in these two zones and when these numbers reach some thresholds, we make the queen evolves by increasing its own size and so increasing the 2 associated zones. After this evolution, the ants have to move some material following the new template function attraction. The low part of the figure shows the evolution of the queen which has evolved 6 times since the simulation beginning. On this figure, we can see the red curves counting the zones density. Each gap in these density curves correspond to an evolution of the queen.



(a) template function



(b) simulation on Repast: after few step



(c) simulation on Repast: after queen adaptive development

Figure 3: Adaptive queen behavior modelling: according to its surround material spatial perception, the queen evolves

MULTI-CRITERIA USER/SERVICE SYSTEM MODELING

To model the concept of multi-criteria phenomena, we introduce different kinds of pheromones. Each kind of pheromone is represented by a specific color. We introduce the notion of center which is a specific spatial location. On each center, we are able to define many queens. Each queen, belonging to a center, is able to emit its own pheromone which is represented by a colored pheromone that is different from the other queen ones belonging to the same center. A queen, associated to a spatial center, describes a specific criterium linked to a colored pheromone. To represent the same criterium on different centers, we describe it by the same colored pheromone on these different spatial locations. In order to force the ants to deposit their material only near the center, we have introduced the template function.

Definition 1 *A spatial multi-criteria multi-center simulation is described by a set of n_p* ***centers****, $\{P_i; 1 \leq i \leq n_p\}$, and by a set of n_c* ***colors****, $\{c_j; 1 \leq j \leq n_c\}$.*
For each center P_i, we define a c_j***-colored template function****, $\Phi_{ij} : S \to I\!R$, which gives the value of the c_j template intensity on each spatial position.*
For each center P_i, we can define a c_j***-colored pheromone function****, $f_{ij} : S \to I\!R$, which gives the value of the c_j pheromone intensity on each spatial position.*

Remark 2 *We can define the c_j-colored template function of the P_i center by the following* **radial exponential function**

$$\Phi_{ij}(x, y) = \alpha_{ij} \exp(-\beta_{ij}(d((x, y), (x_{P_i}, y_{P_i})) - r_{ij})^2) \tag{5}$$

where α_{ij} is the template amplitude, β_{ij} is the template slope, (x_{P_i}, y_{P_i}) are the P_i center coordinates.
We then define the c_j-colored pheromone function for the P_i center with a similar formula

$$f_{ij}(x, y) = a_{ij} \exp(-b_{ij}(d((x, y), (x_{P_i}, y_{P_i})) - r_{ij})^2) \tag{6}$$

where a_{ij} is the template amplitude, b_{ij} is the template slope.

We give in the following some definitions which allow to generalize the ant nest building algorithm for the multi-criteria multi-center simulation.

Definition 3 *A center P_i has the* ***dominant color*** *c_j if*

$$a_{ij} = \max\{a_{ik}; 1 \leq k \leq n_c\}.$$

Definition 4 *On each space location $Z = (x, y)$, we define the c_j* ***colored pheromone intensity*** *as the function $F_j(Z)$ or $F_j(x, y)$ defined by the formula:*

$$F_j(Z) = F_j(x, y) = \sum_{i=1}^{n_p} f_{ij}(x, y). \tag{7}$$

The multi-criteria multi-center model proposed here, implement some spatial objects that are the material and spatial agents (which are the ants). The ants have to carry the material in order to achieve the spatial self-organization simulation.

Definition 5 *A material involved in a spatial multi-criteria multi-center simulation has to include a* ***characteristic color table*** *which corresponds to the only colors that the material is able to perceive and upon which it will be able to react.*

Remark 6 *An ant involved in a spatial multi-criteria multi-center simulation and which is carrying a material has to include a* ***characteristic color table*** *which corresponds to the material characteristic color table.*

Each ant of the simulation which is carrying some material M_l, has to implement a decision process which gives, as output, a color pheromone template c_j that is used for the material transportation by the ant. This selected color c_j is called **the ant behavior**.

At each simulation step, a carried material M_l is associated to a color c_j, called the ant behavior in definition 6. The ant which is carrying this material will then move by searching in its neighboring position, the appropriate one.

Ant behavior computation is based on a specific ranking process which will evaluate the greatest ranking place between the neighboring place corresponding to some highest pheromone color rate.

Definition 7 *For each material M (or the ant carrying it), we define the c_j* ***color attribute preference*** *as the rate, a real number $s_{jM} \in [0, 1]$.*

For each material M (or the ant carrying it) and each space location $Z = (x, y)$, we compute the ***ranking,*** *ρ_{MZ} by the formula:*

$$\rho_{MZ} = \sum_{j=1}^{n_c} s_{jM} * F_j(Z) \tag{8}$$

where n_c is the number of pheromone colors, $F_j(Z)$ is the c_j colored pheromone intensity on the location Z, defined in definition 4 and s_{jM} is the c_j color attribute attribute preference for the material M defined in definition 7.

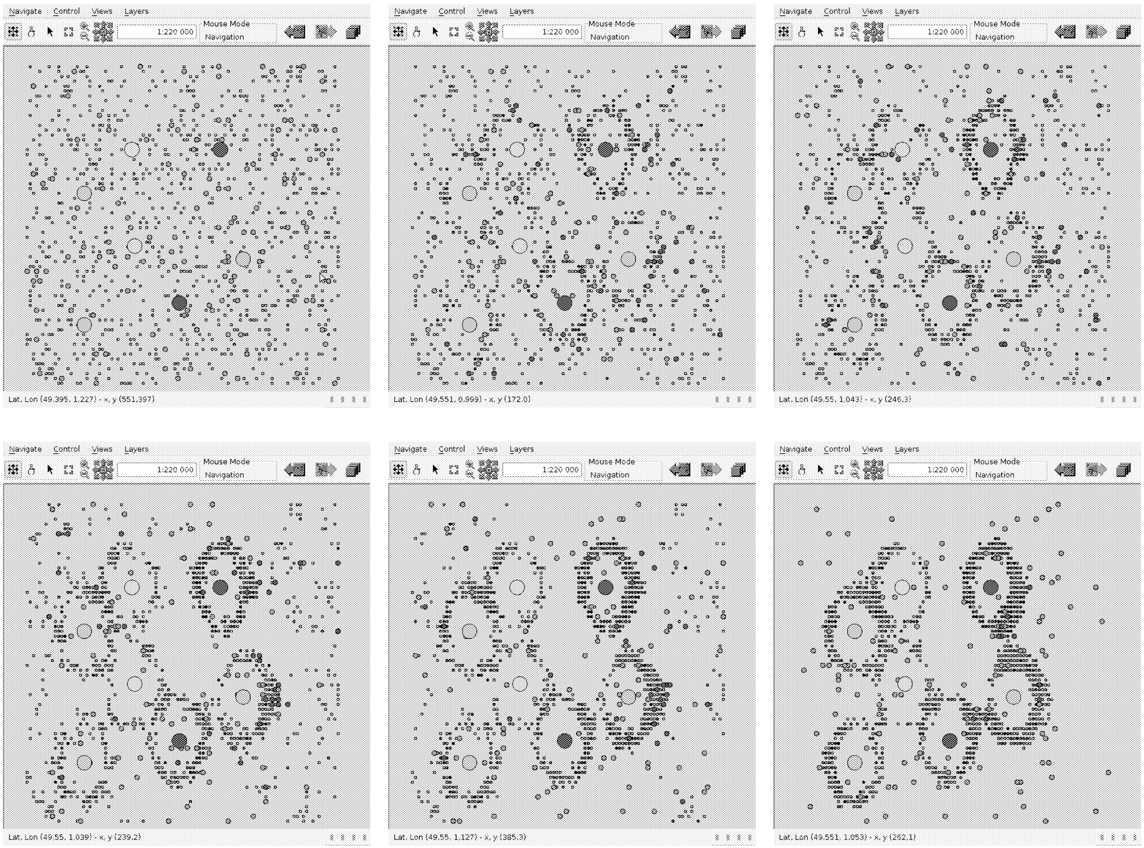


Figure 4: Simulation computation, at successive steps: iterations 0, 152, 250, 370, 601, 1601.

EXPERIMENT OUTPUT AND ANALYSIS

We study an experimental configuration, composed of 6 centers and with random initial places for the materials and for the ants. On each center, we put 8 queens, each one is associated to a colored pheromone labelled from 0 to 7.

On figure 4, we show the result of one simulation where ants progressively aggregate the material around the center, following pheromone trails and clustering algorithm. On the left top sub-figure, we see the initial distribution of materials and ants. In the five other sub-figures, we see three successive steps of the simulation. We can observe the formation of material affectation to each center in order to respect the attraction process, according to the material characteristics.

On figure 5, we make a zoom of the last step of the simulation shown on figure 4, removing the ant representation.

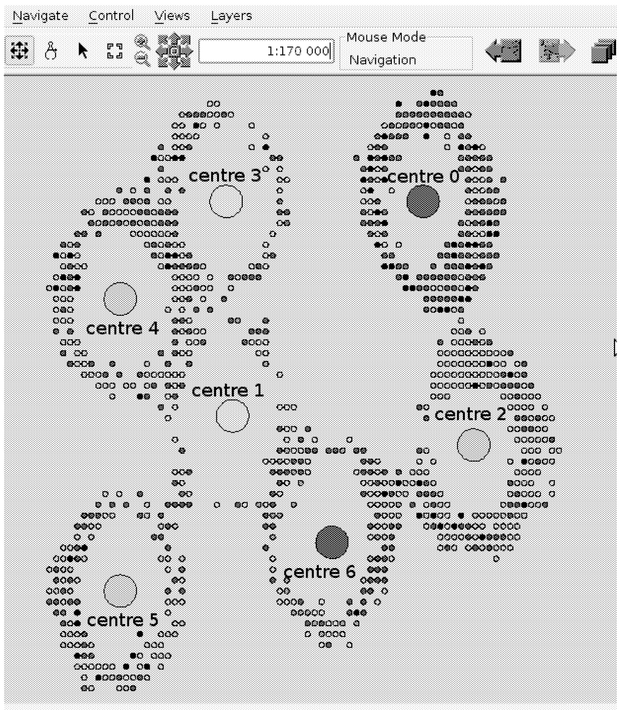


Figure 5: Simulation: zoom on final step

Attraction Analysis Based on Dominant Component

We need to exhibit some analysis to better understand how the computation produced, by self-organization, the distribution of the material over the centers system.

The analysis is based on the dominant component for the material: a characteristic color table is associated to each material. The process of attraction is lead by the ant behavior defined previously. This ant behavior consists here to extract from the characteristic color table, a selected color corresponding to the colored pheromone which controls the ant in order to move to the places of highest values for this colored pheromone.

To better understand the mechanism of attraction, we have to focus on the selected color associated to each material which controls the ant displacement and to forget the other colors belonging to the characteristic color table of the material. In this part, we only associate to each material the dominant color and we study how these colored material are distributed over all the center system.

To analyze this distribution, we represent 3 graphs for each center. On each graph, we have a specific representation of the distribution of all attracted material according to his dominant selected color.

Before defining all these graphs, we have to define a zone of analysis for each center:

Definition 8 *For each center P_i and each color pheromone c_j corresponding to the template function f_{ij} defined in remark 2, we define the* **referential disk** *as*

$$D_{ij} = \{M = (x, y); f_{ij}(M) < ra_{ij}\}$$

where a_{ij} is the template amplitude of the function f_{ij} and $r \in [0, 1]$ is a real number whose value is generally equal to 0.5 *in the following*

According to this referential disk, we compute, by three ways, some indicators corresponding to the quantity of material of each color in this disk, or some relative quantity of material in function of the pheromone amplitude or in function of the neighborhood.

The three graphs used in our study are defined by

- The **material density** of dominant color c_j for the center P_i which is computed as follow:

  $$\rho(P_i, c_j) = \frac{\eta(D_{ij})}{\mathcal{A}(D_{ij})}$$

  where $\eta(D_{ij})$ is the number of materials of dominant color c_j inside the referential disk D_{ij} of the center P_i and $\mathcal{A}(D_{ij})$ is the area (e.g. number of material places) of the disk D_{ij}.

- The **pheromone efficiency** of dominant color c_j for the center P_i which is computed as follow:

  $$\rho_r(P_i, c_j) = \frac{\rho(P_i, c_j)}{a_{ij}}$$

- The **relative pheromone efficiency** graph which computes the queen efficiency relatively to the neighborhood network. This computation consists in changing the pheromone amplitude used in the previous graph by a relative pheromone amplitude a^r_{ij} defined by:

  $$a^r_{ij} = \frac{a_{ij}}{\sum_{k \in \vartheta_i} a_{kj}}$$

  where ϑ_i is the set of centers belonging to the neighborhood of the center c_i on the center neighboring network. The relative pheromone efficiency of dominant color c_j for the center P_i is computed as follow:

  $$\rho_{sr}(P_i, c_j) = \frac{\rho(P_i, c_j)}{a^r_{ij}}$$

The last graph exhibits a complex indicator which takes into account the interaction network of the center system.

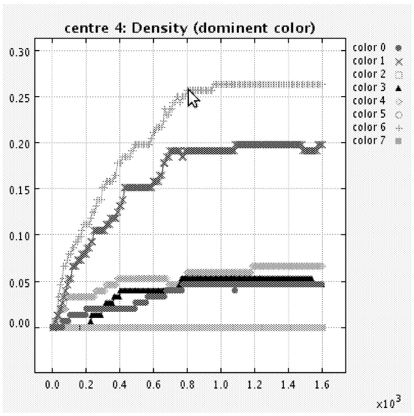
On figure 6, we represent for each center, the 3 graphs defined previously and we represent two additional graphs corresponding to the pheromone amplitude and to the relative pheromone amplitude according to the neighborhood, for each color.

Results Analysis

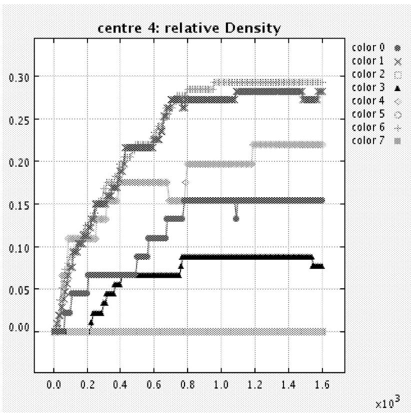
This attraction analysis with the three associated graphs, allows us to better understand the complexity of the phenomena according to the multi-criteria and to the spatial effects. To illustrate this analysis, we will observe the results of the center 4, based on figure 6:

1. A first remark concerns the non linear properties of the attraction phenomenon which makes the color of the more important pheromone intensities attracting a great number of materials of this color and a few number of materials of color of lower pheromone intensities. There is no linear relation between the number of colored material and the corresponding colored pheromone intensity. Finally, the material of color of lower pheromone intensity are not significant. Concerning center 4 of our current experiment, we will focuss only on the material of the two dominant colors: orange (color number 6) and blue (color number 1).

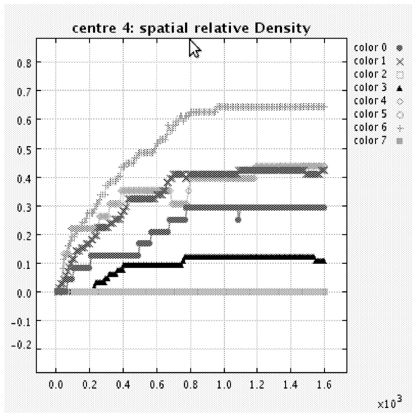
2. On center 4, the first graph - material density - shows that the more attracted colored materials are, in order, the orange colored materials and then



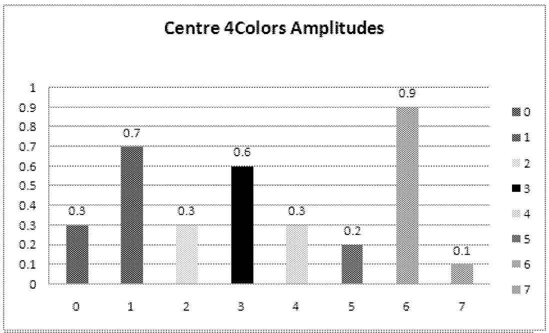
(1) material density



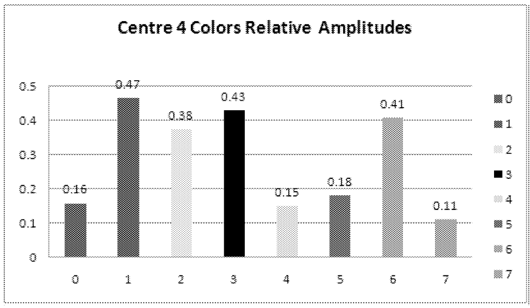
(2) pheromone efficiency



(3) relative pheromone efficiency



(4) pheromone amplitudes



(5) relative pheromone amplitudes

Figure 6: Attraction analysis for the center 4

the blue colored materials. The predominance of the orange colored material over the blue is corrected on the second graph which consists in dividing the colored material number of each color by the corresponding colored pheromone intensity. As orange pheromone is greater than the blue one, we could expect that this center will attract more orange materials than blue ones. The pheromone efficiency graph shows this, making the orange and blue curve become close.

3. The second graph, the pheromone efficiency graph, takes into account only local information about the center and not spatial information. With the third graph, the spatial pheromone efficiency graph, we correct the importance of the pheromone intensity of each color on a specific center with respect to the same color pheromone intensity of the neighboring centers. On figure 5, we observe that the neighbors of center 4 are the centers 1, 3 and 5. For these three centers, the blue pheromone intensity is low and the orange intensity is high on center 5. And so, if we observe part (5) of figure 6, we can see that the relative orange intensity becomes lower than the relative blue pheromone intensity. The last graph - relative pheromone efficiency - gives a correction according to this relative pheromone intensity. But, finally, this graph shows that the orange material number is still greater than the blue one. That is an unpredictable event.

Of course, the complexity of the simulation is not completely predictable by nature and unpredictable phenomena appear as we finally observe in the previous example. There are mainly due to the complexity of the spatial configuration and the multi-criteria characteristics. These unpredictable charateristics of the result overtop our advanced analysis which integrate a spatial interaction correction of first order (e.g. only direct neighbors are considered in this analysis).

CONCLUSION AND PERSPECTIVES

We propose in this work, a general methodology to deal with urban dynamics concerning user/service problems. The model is based on the extension of nest building algorithms to multi-criteria aspects. This extension is based on the management of various elements of the users characteristics which are modeled by specific colors. The simulation presented here consists in implementing attraction phenomena by the use of ranking computations mixed with ant clustering algorithm. Our next perspective is to develop a feed-back process of the center on the user attraction phenomenon. This feed-back process is under development and consists in analysing the result of attraction phenomena with the method proposed in this paper. Rule-based systems using data from this analysis for specific centers will lead to modify some of their colored pheromone rate. The whole processus of attraction and feed-back will be described in this model allowing to better understand the complexity of the spatial dynamics and its consequence for decision making for urban planning.

REFERENCES

[1] Bertelle,C.; Duchamp, G.H.E. and H. Kadri-Dahmani (eds) (2008) *Complex Systems and Self-Organization Modelling*, "Understanding Complex Systems" series, Springer, (in press).

[2] Bonabeau, E.; Dorigo, M. and G. Theraulaz (1999) *Swam Intelligence, from natural to artificial systems*, "Santa Fe Institute Studies in the Sciences of Complexity" series, Oxford University Press.

[3] Camazine, S.; Deneubourg, J.-L.; Franks, N.R.; Sneyd, J.; Theraulaz, G. and E. Bonabeau (2001) *Self-Organization in Biological Systems*, Princeton University Press.

[4] Rawan Ghnemat (2009) *Adaptive Modeling for Spatial Emergence within Complex Systems*, PhD Thesis, Le Havre University, France.

[5] Ghnemat, R.; Bertelle, C. and G.H.E Duchamp (2007) *Adaptive Automata Community Detection and Clustering, a generic methodology*, in *Proceedings of World Congress on Engineering 2007, International Conference of Computational Intelligence and Intelligent Systems*, pp 25-30, London, U.K., 2-4th July 2007.

[6] Repast web site (20089) `http://repast.sourceforge.net`.

AUTHOR LISTING

AUTHOR LISTING