Journal Subline

LNCS 7220

Maiga Chang • Mingmin Zhang
Guest Editors

Transactions on Edutainment VIII

Zhigeng Pan • Adrian David Cheok • Wolfgang Müller
Editors-in-Chief

Springer

Lecture Notes in Computer Science 7220

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison
Lancaster University, UK

Takeo Kanade
Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler
University of Surrey, Guildford, UK

Jon M. Kleinberg
Cornell University, Ithaca, NY, USA

Friedemann Mattern
ETH Zurich, Switzerland

John C. Mitchell
Stanford University, CA, USA

Moni Naor
Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz
University of Bern, Switzerland

C. Pandu Rangan
Indian Institute of Technology, Madras, India

Bernhard Steffen
TU Dortmund University, Germany

Madhu Sudan
Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos
University of California, Los Angeles, CA, USA

Doug Tygar
University of California, Berkeley, CA, USA

Moshe Y. Vardi
Rice University, Houston, TX, USA

Gerhard Weikum
Max Planck Institute for Informatics, Saarbruecken, Germany

Zhigeng Pan Adrian David Cheok
Wolfgang Müller Maiga Chang
Mingmin Zhang (Eds.)

Transactions on Edutainment VIII

Springer

Editors-in-Chief

Zhigeng Pan
Hangzhou Normal University, Hangzhou, China,
E-mail: zhigengpan@gmail.com

Adrian David Cheok
National University of Singapore, Singapore
E-mail: adriancheok@mixedrealitylab.org

Wolfgang Müller
University of Education, Weingarten, Germany
E-mail: mueller@md-phw.de

Guest Editors

Maiga Chang
Athabasca University, AB, Canada
E-mail: maigac@athabascau.ca

Mingmin Zhang
Zhejiang University, Hangzhou, China
E-mail: zmm@cad.zju.edu.cn

ISSN 0302-9743 (LNCS)
ISSN 1867-7207 (TEDUTAIN)
ISBN 978-3-642-31438-4
DOI 10.1007/978-3-642-31439-1
Springer Heidelberg Dordrecht London New York

e-ISSN 1611-3349 (LNCS)
e-ISSN 1867-7754 (TEDUTAIN)
e-ISBN 978-3-642-31439-1

CR Subject Classification (1998): K.3.1-2, H.5.1-3, H.5, K.3, I.3-5, H.3, I.2.1

© Springer-Verlag Berlin Heidelberg 2012
This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in ist current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.
The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

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

Editorial

In this journal issue, we have 24 papers from different sources. The first ten articles of this issue represent a selection of outstanding contributions from Edutainment 2011, the 6th International Conference on E-Learning and Games held in Taiwan, in September 2011. The main purpose of Edutainment conferences is the discussion, presentation, and information exchange of scientific and technological developments in the new community. These ten papers cover a broad area from game engines, using games to teach, identifying a player's emotional states, and assessing the effect of educational games on multi-touch interaction, natural user interface, and virtual reality: "Analysis of Brainwave Characteristics for Playing Heterogeneous Computer Games;" "Exploitation in Context-Sensitive Affect Sensing from Improvisational Interaction;" "Hybrid Document Matching Method for Page Identification of Digilog Books;" "Game Design Considerations when Using Non-Touch-Based Natural User Interface;" "Effects of Game-Based Learning on Novice's Motivation, Flow, and Performance;" "Behavioral Traits of the Online Parent–Child Game Players: A Case Study and Its Inspirations;" "Towards an Open Source Game Engine for Teaching and Research;" "Designing a Mixed Digital Signage and Multi-touch Interaction for Social Learning;" "Old Dogs Can Learn New Tricks: Exploring Effective Strategies to Facilitate Somatosensory Video Games for Institutionalized Older Veterans;" "Effects of Multi-symbols on Enhancing Virtual Reality Based Collaborative."

The other 14 papers are regular papers or were selected from national conferences. In the paper "Particle Filter with Affine Transformation for Multiple Key Points Tracking," Sheng Liu et al. propose a novel multiple-point tracking method for long microscopic sequences, which is tested on the micro stereo imaging system and demonstrated by the experimental results. In the paper "Meshless Simulation of Plastic Deformation and Crack," Liang Zeng et al. present a meshless method for local deformation and crack simulation of plastic thin shells. In "Lossless Visible Three-Dimensional Watermark of Digital Elevation Model Data," Yong Luo et al. present a lossless visible 3D watermarking algorithm for protecting digital elevation model data; experiments demonstrate that the proposed algorithm has satisfactory security and can effectively protect the copyright of digital elevation model data. In the paper "3D Articulated Hand Tracking Based On Behavioral Model," Zhiquan Feng et al. propose a new cognitive and behavioral model, TPTM model, providing a way to connect users and computers for effective interaction, and present a real-time freehand tracking algorithm that tracks 3D freehand in real-time with high accuracy. In the paper "SPIDAR-Pen: A 2T1R Pen-Based Interface with Co-located Haptic-Visual Display," the authors propose a string-driven pen-based interface, which is capable of two translational and one rotational (2T1R) interactions, and co-located haptic-visual display. In the paper "Generation of IFS Fractal Images

Based on Hidden Markov Model," Liliang Zhang et al. present a method of generation of iterated function systems fractal attractor images based on a hidden Markov model. The method can draw the shape and color of fractal images by using a probability matrix. In the paper "High Performance Hybrid Streaming Scheme and VCR Support," Yong Wang et al. propose a new scheme, HPHS, and discuss how to support interactive VCR. It merges the outstanding thoughts of HPSCT (revised from GEBB) and a patching scheme. The scheme aims at delivering popular video, allocates the number of channels adaptively according to the request rate to minimize the network bandwidth consumption and satisfy the real-time playback requirement. In the paper "A Modular Image Search Engine Based on Key Words and Color Features," Xianying Huang et al. presents a modular image search engine based on keywords and contents, which organically combines the search engine technology of keywords and color feature of the images; a quantified method based on the maximum pixels ratio of irregular connected regions is also raised, and the retrieval efficiency is proved by experiments. In the paper "The Corroded Defect Rating System of Coating Material Based on Computer Vision," Gang Ji et al. present a method based on computer vision to solve the problem of artificial accurate detection of coating material corroded defects and making accurate ratings automatically. In the paper "Service Replacement Algorithm Using a Complex Network," Yang Zhang et al. use the replaceable degree from the dependency relationship to estimate the effect of the depth and breadth of dependencies between services on the replacement process and its results. In the paper "A Virtual Training System Using a Force Feedback Haptic Device for Oral Implantology," Xiaojun Chen et al. present a virtual training system, by which the resulting data of the preoperative planning can be transferred, and surgical simulation of the plan can be vividly realized. In the paper, "Modeling and Analysis of the Business Process of the Supply Chain of Virtual Enterprises" by Min Lu et al, UML and object-oriented Petri nets are combined to model the distributed business process of the supply chain of virtual enterprises. In the paper "Tree Branching Reconstruction from Unilateral Point Clouds," Yinghui Wang et al. perform unilateral scanning of real-world trees and propose an approach that could reconstruct trees from incomplete point clouds. In the final paper, "Literature Analysis on the Higher Education Informatization of China (2001–2010)," Chen et al. present statistical results from four dimensions: basic theory, construction, management, and evaluation of higher education informatization in China. The authors analyze the statistical results from the distribution of the number of articles, authors, and contents of the 278 articles.

The papers in this issue present numerous application examples of edutainment, which gives more evidence of the high potential and impact of edutainment approaches. We would like to express our thanks to all those people who contributed to this issue. The authors of the papers, the reviewers, and the Program Committee of Edutainment 2011 for recommending high-quality papers to this

new journal. Special thanks go to Fotis Liarokapis, Ming-Puu Chen, Wu-Yuin Hwang, and Wolfgang Müller, who helped select these papers from the two conferences. In addition, thanks also go to Yi Li, Fan Dai, and Qiaoyun Chen from the Editorial Office of this journal in Nanjing Normal University—they spent a lot of effort in contacting authors, managing the review process, checking the format of all the papers and collecting all the material.

April 2012

Maiga Chang
Mingmin Zhang

Transactions on Edutainment

This journal subline serves as a forum for stimulating and disseminating innovative research ideas, theories, emerging technologies, empirical investigations, state-of-the-art methods, and tools in all different genres of edutainment, such as game-based learning and serious games, interactive storytelling, virtual learning environments, VR-based education, and related fields. It covers aspects from educational and game theories, human-computer interaction, computer graphics, artificial intelligence, and systems design.

Editors-in-Chief

Zhigeng Pan	Hangzhou Normal University, China
Adrian David Cheok	NUS, Singapore
Wolfgang Müller	University of Education Weingarten, Germany

Managing Editor

Yi Li	Nanjing Normal University, China

Editorial Board

Ruth Aylett	Heriot-Watt University, UK
Judith Brown	Brown Cunningham Associates, USA
Yiyu Cai	NTU, Singapore
Maiga Chang	Athabasca University, Canada
Holger Diener	Fhg-IGD Rostock, Germany
Jayfus Tucker Doswell	Juxtopia Group, USA
Sara de Freitas	The Serious Games Institute, UK
Lynne Hall	University of Sunderland, UK
Masa Inakage	Keio University, Japan
Ido A. Iurgel	Universidade do Minho, Portugal
Kárpáti Andrea	Eötvös Loránd University, Hungary
Lars Kjelldahl	KTH, Sweden
James Lester	North Carolina State University, USA
Nicolas Mollet	IIT, Italy
Ryohei Nakatsu	NUS, Singapore
Ana Paiva	INESC-ID, Portugal
Abdennour El Rhalibi	JMU, UK
Daniel Thalmann	EPFL, Switzerland

Kok-Wai Wong Murdoch University, Australia
Gangshan Wu Nanjing University, China
Xiaopeng Zhang IA-CAS, China
Stefan Goebel ZGDV, Germany
Michitaka Hirose University of Tokyo, Japan
Hyun Seung Yang KAIST, Korea

Editorial Assistants

Ruwei Yun Nanjing Normal University, China
Qiaoyun Chen Nanjing Normal University, China

Editorial Office

Address: Ninghai Road 122, Edu-Game Research Center, School of Education Science, Nanjing Normal University, Nanjing, 210097, China
E-mail: njnu.edutainment@gmail.com; edutainment@njnu.edu.cn
Tel/Fax: 86-25-83598921

Table of Contents

Papers from Edutainment 2011

Regular Papers

Analysis of Brainwave Characteristics for Playing Heterogeneous Computer Games

Fu-Chien Kao, Han-Chien Hsieh, and Wei-Te Li

Computer Science & Information Engineering, Dayeh University, Taiwan
{fuchien,R9906020,R9906031}@mail.dyu.edu.tw

Abstract. The traditional E-learning often offers the online examination to assess the learning effect of a student after completion of the online learning. Basically, this traditional learning assessment mechanism is a passive and negative assessment mechanism, which cannot provide an real-time learning warning mechanism for teachers or students to find out problems as early as possible (including such learning conditions as "absence of mind" resulting from poor learning stage or physical or psychological factor), and the post-assessment mechanism also cannot assess the learning effectiveness provided by the online learning system. This research paper attempts to acquire the electroencephalogram to analyze the characteristic frequency band of the brainwave related to learning and formulate the learning energy index (LEI) for the learner at the time when the learner is reasoning logically via the brain-wave detector based on the cognitive neuroscience. With the established LEI, the physical and psychological conditions of an online leaner can be provided instantly for teachers for assessment. Given that the learning system is integrated into the brainwave analytic sensing component, the system not only can provide learners an instant learning warming mechanism, but also help teachers and learning partners to further understand the causes of learning disorder of learners, and can also provide relevant learning members with timely care and encouragement. Besides, this research also would prove that the game-based learning has not only the energy distribution of the characteristic frequency band the same as that by using professional textbooks, but also the way of game design can enhance the LEI of learners more in the aspect of training logical reasoning.

Keywords: Cognitive neuroscience, Learning energy index, Electroencephalogram, Game-based learning.

1 Introduction

The cognitive neuroscience is mainly to explore the relationship between brain and noema. In fact, it is difficult to make a distinction among cognitive neuroscience, cognitive science and neuroscience. The cognitive neuroscience based on the neuroscience uses the special structure and signal of the neural system to interpret its cognitive function. In other word, it, which is to understand how the nerve cells transmit messages and communicate with each other from the view of biology when we

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 1–11, 2012.
© Springer-Verlag Berlin Heidelberg 2012

study the brain system and cognition, is a functional neuroscience on the basis of cognitive science and neuroscience [1]. The cognitive science, which is a concept containing wide meaning, covers the sociology, the psychology, the pedagogy, artificial intelligence, linguistics, bioengineering, etc. and therefore is an interdisciplinary science [2]. The research range of cognitive science includes the method of storing and displaying messages and knowledge in brain, the way of human understanding languages and expressing thoughts and feelings by languages, the way of understanding images and generating consciousness, thoughts and feelings, the way of learning new knowledge, reasoning and judging, solving problems, establishing and implementing plans, which are the psychological activity caused by the intelligent mechanism of brain.

Cognition means that the noema development of individuals is closely related to knowledge, learning and intelligence, and must be transmitted via brain and mental activities. The cognitive psychologists emphasized that learning is a process of initiatively receiving messages and creative thinking, and found that the formation of human cognition could be considered as a process of processing data by computer. They attempted to use the human brain to simulate the operation flow of computer, view individuals as a data processing system, which can actively select messages generated by environment and can process and store further through mental activities so as to be ready for information retrieval and extraction when required [3-4].

Consisting of neuron, colloid and other sustenticular cells, the cerebrum suspends in the cerebrospinal fluid of the cranial cavity and is the most stringent guarded organ. The cerebrum is also considered as the central nervous system, which is the sensorium controlling all thinking and receives messages generated by the external environment so as to give out commands related to various activities and external behaviors through the integration of the neural system of the cerebrum and the complex thoughts and actions in the brain. In recent years, with the rapid development of the medical technologies and the improvement of the detection instruments, all parts of the cerebrum and their relevant functions have been further known. Through observation of the function and the sensory perception of the brain, we can know that every part of the brain controls a different task. The relevant functional parts of the brain contain the parietal lobe, the occipital lobe, the temporal lobe and the frontal lobe, of which the parietal lobe is located at the top of the head and the recognition center of the algesia and the feeling, the occipital lobe is located at the afterbrain and responsible for the visual recognition and related to the abilities of word recognition and geometry learning, and the frontal lobe is located at the forehead and responsible for motion coordination, thinking and judgment and problem solving and is the commander in chief of the brain.

Based on the neuroscience, the cognitive neuroscience uses the network structure and signal transmission of the neural system to interpret the cognitive function. In other word, it is a functional neuroscience for interneuronal information transmission and communication when the cognitive behavior of the brain is researched. With the development of the neuroscience, technologies and instruments for research of human brain have been significantly improved in recent years. Scientists use advanced instruments such as magnetic resonance imaging (MRI), functional magnetic resonance imaging (FMRI), computerized tomography (CT), positron emission

tomography (PET), magneto encephalography (MEG), electroencephalogram (EEG), etc. to explore the cerebral activities so that researchers can further know the relationship between brain and mental activities [5] when human is carrying out a special action. The application of the electroencephalogram in the human body is approximately 40 years earlier than the MEG. The electroencephalogram is mainly used for recording the potential difference of the brain, and the MEG is used for recording the magnetic field of the brain. Both of them are often applied for exploration of the potential excitement source of the brain. The electroencephalogram is characterized by noninvasive measurement, non-radiation, higher time resolution, transplantability and low cost, so this research adopts the electroencephalogram to study its characteristic frequency band.

2 Electroencephalogram Measurement

The electroencephalogram is the electromagnetic signal generated by the neural activities of the brain, which is the message transmitted by the cranial nerve using the electrochemical process. The nerves convert the chemical signal into the electric signal through release of the chemical substance. These electric signals are mainly generated by the ion permeability. The neurons consisting of nerves are cells specially for transmitting electric signal. The neurons will produce the action potential to transmit message when the receptors on the cell surface have received the neurotransmitters. Different stimulations will have different transmission signals. Rhythmic potential changes will occur when the nerve impulse is transmitted between the nerves and the nerve fibers. These weak potential changes consist of the electrical rhythms and the transient discharge. It is the so-called electroencephalogram when these neurons generate the resultant potential changes.

The energy of electroencephalogram action is far less than other psychological signals of the human body (e.g. electrocardiogram and electromyogram), so it is difficult to record the electroencephalogram. The main reason is that its signal is so weak that the measurable brainwave is about 1 to 100μV and the frequency range is from 0.5 to 100Hz [6] when the brainwave signals pass through the meninges, the cerebrospinal fluid and skull and reach the scalp. The electroencephalogram basically has the synchronization, and will be affected by different waking states and also have different characteristics in different sleeping periods.

There are four frequency bands such as Alpha (α), Beta (β), Theta (θ) and Delta (δ) [6] according to the data provided by the International Federation of Societies for Electroencephalography and Clinical Neurophysiology. The α wave is a brainwave with frequency between 8Hz and 13Hz, and with amplitude of about 50μV. The β wave has a main frequency between 13Hz and 22 Hz and a potential of about 5-20 μV. In the case of keeping a clear mind and alert, it often appears in the parietal region and the frontal region. And this wave will be more obvious especially when the brain is in the case of logical thinking, computation, reasoning or sensory stimulus. The θ wave has a frequency between 4Hz and 8 Hz. Generally speaking, its amplitude is lower. This wave often appears in the parietal region and temporal region of the

children, and will occur when the emotion of adult is suppressed or the consciousness of a person is interrupted, the body is relaxed or in the case of meditation. The δ wave has a frequency between 0.2Hz and 4Hz. It will become more obvious when people fall asleep deeply or babies fall asleep. For relevant functional parts of the brain, refer to figure 1.

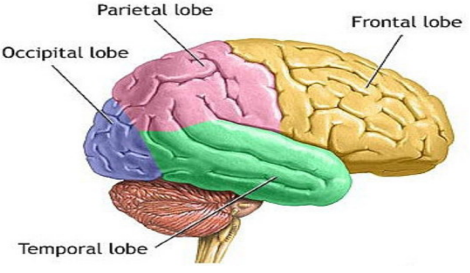


Fig. 1. The four functional areas of the brain

The brainwave measurement is basically divided into the monopolar type, the average type and the bipolar type. The position of the electrodes for the brainwave measurement has a direct influence on amplitude, phase and frequency of the brainwave signal, so the brainwave signals measured at different electrode positions cannot be compared with each other. The International Federation of Societies for Electroencephalography and Clinical Neurophysiology established a set of 10-20 electrode system. The top view of the relevant electrode positions are as shown in figure 2 [8]. Each electrode point is expressed by the combination of letters and digits. The first letter A represents the earlobe, C represents the central sulcus of the frontal lobe, P represents the parietal lobe, F represents the frontal lobe, T represents the temporal lobe and O represents the occipital lobe. If the second place is an even number, it is in the right cerebrum, and if an odd number, it is in the left cerebrum. If the second place is the letter Z, it means the electrode point is located on the central line of the nasion to the inion. Fp1 and Fp2 are respectively located in the left and right frontal lobe.This research adopts a simple there-electrode measurement to carry out the application study of relevant brainwaves based on the considerations of the transplantability and development, etc.

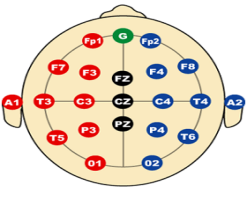


Fig. 2. T-20 standard electrode International Location

3 Design of EEG Measuring Module

With the continuous improvement of material technology and measuring technology, the measuring technology of the brainwave has been accelerated in recent years. A medical grade electroencephalograph, however, is huge, expensive and complicated in processing signals of the backend. Therefore, it is required that the operation shall be carried out by professional staff during the measurement. The brainwave measuring module proposed in this research is shown as in figure 3, which is not only small in size, convenient to carry and easy to operate but also is low in price, and is applicable to being used in various industries in the future compared to the medical grade electroencephalograph. This research uses a simple bipolar system to measure the brainwave. The relevant electrode positions are respectively in the forehead and the lower earlobe and used for recording signals at Fp1, Fp2, A1, etc. of which the A1 is the reference potential. The measuring positions are as shown in figure 4.

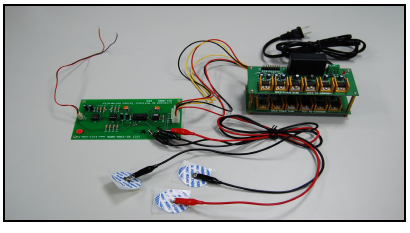


Fig. 3. EEG measurement module design

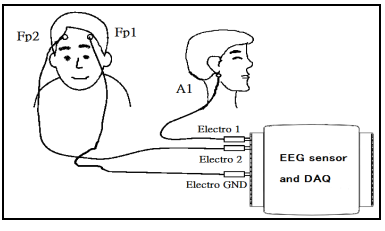


Fig. 4. Measuring schematic

3.1 Design of Brainwave Measuring Circuit

During measurement, hair on the head will prevent the electrodes from being stuck, so it is difficult for the participants to stick the electrodes on the scalp. Therefore, the measurement of electrodes is carried out at positions of the forehead Fp1 and Fp2 and the lower earlobe A1. A1 is used as a reference potential. An isolation measure shall be taken so as to avoid electric shock due to the electric leakage of the power supply or instrument. The block diagram of relevant brainwave measuring circuit and the brainwave sensor board are described in figure 5 and figure 6. The preamplifier uses the instrument amplifier to capture the vector signal of the electroencephalogram as the monopolar signal with a magnification of 50, and uses the JFET operational amplifier to increase the impedance match between electrodes and circuit. The bandwidth of the band-pass filter is set as 1Hz to 20Hz, of which the purpose is to screen out the four frequency-band signals such as α, β, δ and θ, and the signals passed through the filter are amplified to 1000 times gain.

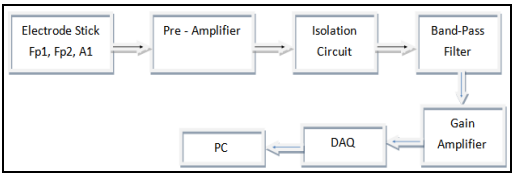


Fig. 5. EEG functional block diagram of measuring circuit

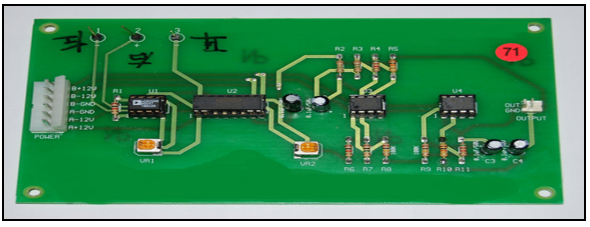


Fig. 6. EEG measurement circuits

4 System Implementation and Statistic Analysis

This research establishes a LEI from the point of view of the cognitive neuroscience. This LEI is expected to provide the teachers and learners with the capabilities of observing the physiological condition of learners or the learning members of the same group if they are appropriate to the online learning so as to avoid the absence of mind and to improve effective the learning results when the future online learning cooperation is conducted. Furthermore, the establishment of the LEI also needs the analysis of the brainwave characteristic frequency band. The relevant analysis of the brainwave characteristic frequency band can also provide you with a basis for R&D of the hardware of the quick-detecting module for the brainwave energy in the future learning. In this research, not only the relevant brainwave-measuring modules and the brainwave characteristic analysis programs are actually made, but also discuss the following subjects: (1) Analysis of the brainwave characteristic frequency band for human under different physiological statuses (falling asleep, deep sleep, logical reasoning, reading, etc.); (2) Comparison analysis of brainwave difference when reading the print textbooks and the multi-medium textbooks; (3) Whether habit formation is helpful to learning; (4) Whether game-base learning is the positive learning.

4.1 Framework of Analysis System

The system framework proposed in this research includes the front brainwave-sensing module for capturing brainwave signals and the rear brainwave-analysis program for analyzing the characteristics of brainwave frequency band. After the brainwave signals of the participants are acquired by the front brainwave-sensing module, the digitalized brainwave signals will be sent a computer and saved as a specified format and provided the rear brainwave-analysis program for the characteristics analysis of brainwave frequency band through the analog to digital converter (ADC) of the DAQC. The relevant framework for characteristics analysis of the brainwave frequency band is as shown in figure 7. The relevant steps are described as follows:

(1) After installation of a brainwave sensor, the electrode patches are attached to the participants and then the LabVIEW acquisition program is used to capture brainwave signals.

(2) After being converted by the ADC module of the sensor, the brainwave signals are sent to PC and saved as Excel or Txt format through the USB port of the DAQC.
(3) The EEG Analysis GUI provides brainwave analysis for the data in the format of Excel or Txt. The time-domain part of the GUI provides the strength change in time for the original brainwave signals. After the brainwave signals have been processed by FFT formally. In this research, the percentage of amplitude to energy of main brain frequency band is used to calculate the LEI.

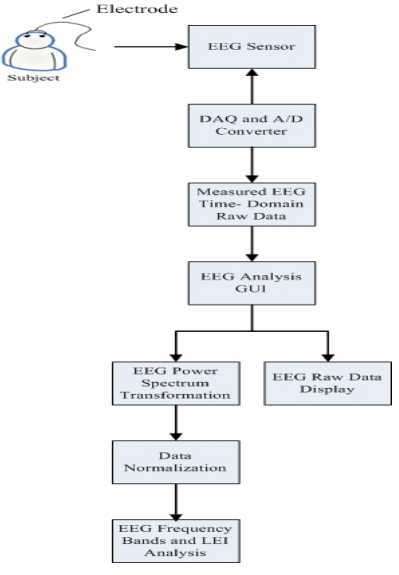


Fig. 7. Analysis of EEG frequency band features of the system architecture diagram

4.2 Method of Brainwave Signal Analysis

In this research, a brainwave analysis system with GUI was developed by the MATLAB program language. The system uses the fast Fourier transform (FFT) to convert the time-domain signals into the corresponding frequency spectrum signals. The analysis interface is shown as in figure 8.

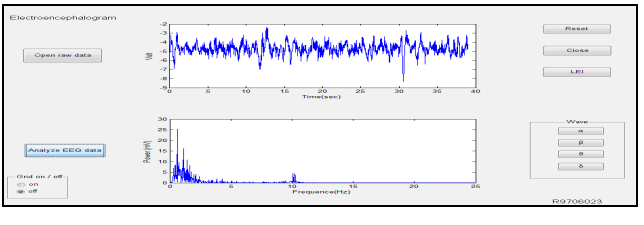


Fig. 8. EEG analysis interface

4.3 Learning Energy Index (LEI)

In this research, 50 students from the Department of Computer Science were tested for their brainwave signals for the logical reasoning of professional books. Based on the energy distribution of main brainwave characteristic frequency band by analyzing learners' logical reasoning ability, it establishes the LEI. The energy distribution of the tested brainwave characteristic frequency band is listed in Table 1. Based on the α, β, θ and δ in Table 1, the average energy percentage of the frequency band of each zone is calculated and the relevant characteristic frequency bands are found out according to the energy level when the participants were tested for their "Probability and Statistics" of the professional discipline. The test results showed that 46 students whose subzone characteristic frequencies of the α and β apparently increased in energy, the frequency of relevant subzones for α is 9Hz to 11Hz and for β is 19Hz to 21Hz, the main frequencies of the subzones for α are α2 and α3 and for β are β7 and β8, and the energy percentage of relevant subzones is apparently higher than that of other subzones when the 50 students were reasoning logically in a state of consciousness (see Table 1). From the Table1, we can know the characteristic frequency bands of the subzone are the frequency bands at δ1, δ2, θ1 and θ2 at the time when people fall asleep or are in deep sleep, which are apparently different from those generated by logical reasoning in the case of consciousness, and the energy of the frequency bands of α and β is obviously lower than that acquired in the case of consciousness. Therefore, in order to consider the brainwave energy of the quick-detecting learning, the total energy percentage related to frequencies of main zones (α2 and α3, and β7 and β8) is specified as the LEI, and the LEI is used as the basis for teachers to understand the physiological status of studying alone or cooperatively of learners.

The average of the total potential amplitude of different frequency bands for 50 participants is calculated so as to obtain the energy of the zone frequency band and the total energy using equations (1) and (2).

$$E_B = \sum_{f}^{n} Power_f \tag{1}$$

$$E_T = \sum_{f=0.2}^{25} Power_f \tag{2}$$

In the above equations, B are the zone frequency bands, f is the start frequency of each frequency band, n is the end frequency of each frequency band (the frequency sampling interval is 0.01Hz), and E is energy of each frequency band. E_T is the total energy of the four zone frequency bands from 0.2Hz to 25Hz. The energy percentage of α, β, θ and δ is respectively (E_B/E_T) %. The energy percentage of the subzone frequency E_Δ is namely the percentage of the energy in the individual subzone and the energy in the total frequency band, as shown in equation (3):

$$E_\Delta = \frac{E_\Delta}{E_T}(\%) \tag{3}$$

Table 1. Energy distribution of frequency band for the learning brainwave

Status	Brain wave type	Subzone frequency(Hz)		Subzone energy/ total energy (E_Δ%)	Total energy percentage of zone (E_B/E_T %)	Characteristic frequency of subzone
Learning	Delta (δ) 0.2 ~ 4 Hz	δ1	0.2 ~ 1	2.38	8.11	
		δ2	1 ~ 2	1.93		
		δ3	2 ~ 3	1.77		
		δ4	3 ~ 4	2.03		
	Theta (θ) 4 ~ 8 Hz	θ1	4 ~ 5	1.95	9.34	
		θ2	5 ~ 6	2.05		
		θ3	6 ~ 7	2.40		
		θ4	7 ~ 8	2.93		
	Alpha (α) 8 ~ 13 Hz	α1	8 ~ 9	3.27	43.23	**α2** **α3**
		α2	9 ~ 10	12.92		
		α3	10 ~ 11	20.93		
		α4	11 ~ 12	3.56		
		α5	12 ~ 13	2.55		
	Beta (β) 13 ~ 22Hz	β1	13 ~ 14	3.44	39.32	**β7** **β8**
		β2	14 ~ 15	2.88		
		β3	15 ~ 16	3.03		
		β4	16 ~ 17	2.97		
		β5	17 ~ 18	2.72		
		β6	18 ~ 19	3.09		
		β7	19 ~ 20	12.70		
		β8	20 ~ 21	5.64		
		β9	21 ~ 22	2.84		
LEI(%)		**$LEI = \alpha_{2-3} + \beta_{7-8} = 52.19$, $\alpha_{2-3} = 33.85$, $\beta_{7-8} = 18.34$**				

4.4 Analysis of Brainwave Characteristics for Playing Heterogeneous Computer Games

With the progress of computer equipment and the popularization of the network, digital games have been gradually a part of leisure and entertainment for modern people, and even played an indispensable role in work in recent years. Some educationists think that the design of game has its attraction for people and thus suggest the game-base learning to improve the learning effect. The public, however,

has a jaundiced view of the negative effects on schoolwork of students caused by the general computer multi-medium game. This research, therefore, is about to discuss if the relevant computer games are helpful to the positive learning from the point of view of the cognitive neuroscience. In this game experiment, a total of 30 people are tested. In this research, there are three games, which are respectively the puzzles, web games and action games, of which the puzzles use the multi-medium Sudoku, the web games use the Happy Farm constructed on the network community "Facebook", and the action games are shooting games with effects of sound, light and 3D.

Participants whose average electroencephalograms are as shown in figure 9 when they are playing Sudoku game. The observations of relevant electroencephalograms showed that there is no big difference between the peak amplitudes of α and β when playing Sudoku, in which the participants carry out the logical reasoning in the case of keeping clear-headed and relaxed; the peak amplitude of α is higher than that of β when playing Happy Farm, which means that the participants need no much thinking and are in a relaxed environment; the peak amplitude of β is higher than that of α when playing the shooting game, which means the participants are nervous and need to make decisions for shooting the target object at any time. The LEI and the energy distribution of the brainwave characteristic frequency band are listed in Table 2 for the 3 heterogeneous games, of which the LEI values approximate 79% and the characteristic frequency bands are respectively located on 9Hz to 11Hz, 19Hz to 21Hz, etc., the same as those measured in the professional science learning. We can know, therefore, the computer games can also provide the logical thinking and decision-making learning with positive help.

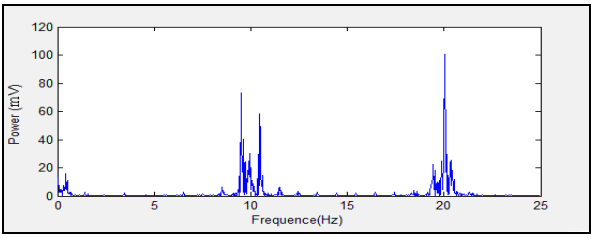


Fig. 9. Averaged brain wave when playing Sudoku puzzle game

Table 2. Energy distribution of characteristic frequency band for heterogeneous games

Type of multi-medium games	LEI (%)	Characteristic frequency band α2–3(%)	Characteristic frequency band β7–8(%)
Sudoku	78.55	39.64	38.91
Happy Farm	79.86	43.12	36.74
Shooting game	79.06	36.92	42.14

5 Conclusion

In this research, the graphic processing software tool LabVIEW and the program development software MATLAB are used to develop a brainwave analysis system for analyzing the brainwave energy of frequency band of learners. Based on the analysis of the acquired brainwave signals and the brainwave energy through the brainwave sensor, the LEI for learners is established for observing the learners or learning members of the same group whose physiological status to see if they are appropriate to the online learning so as to avoid the absence of mind and to improve effective the learning results when the future online learning cooperation is conducted. Furthermore, the establishment of the learning energy index also needs analysis of the characteristic frequency band of the brainwave. The relevant characteristic frequency band of the brainwave can also be the basis for the quick-detecting module for the brainwave energy in the future learning. In this research, various frequency spectrum properties of brainwave are not only analyzed and discussed for learners conducting diverse learning, but also the brainwave difference caused by using the traditional print textbooks and the multi-medium computer games from the view of the cognitive neuroscience. From the analysis of the experimental results, it is to prove that the game-based learning has not only the energy distribution of the characteristic frequency band the same as that by using professional textbooks, but also the way of game design can enhance the LEI of learners more in the aspect of training logical reasoning.

References

1. Albright, T.D., Neville, H.J.: Neurosciences. MITECS: li-lxxii (1999), http://cognet.mit.edu/MITECS/Articles/neurointro.html
2. Eysenck, M.W., Keane, M.T.: Cognitive Psychology: A Student's Handbook
3. Atkinson, R.C., Shffirn, R.M.: Human memory: A propowed system and its control processes. In: Spence, K.W., Spence, J.T. (eds.) Advances in the Psychology of Learning and Motivation Research and Theory, vol. 2, Academic Press, New York (1968)
4. Gagne, R.M.: The condition of learning, 4th edn. Holt Rinehart & Winston, New York (1985)
5. Jensen, E.P.: Brain-based learning: The new paradigm of teaching. Corwin Press, Thousand Oaks (2008)
6. Webster, J.G.: Electroencephalography: Brain electrical activity. Encyclopedia of Medical Devices and Instrumentation 2, 1084–1107 (1988)
7. Hu, M.M.: Physiology. Ho-Chi Book Publishing Co. (1991)
8. Schaul, N.: The Fundamental Neural Mechanisms of Electroencephalography. Electroencephalography and Clinical Neurophysiology 106, 101–107 (1998)
9. Guan, S.Y.: EEG Explanation. Yi-Hsien Publishing Co. (2002)

Exploitation in Context-Sensitive Affect Sensing from Improvisational Interaction

Li Zhang

School of Computing, Engineering and Information Sciences, Northumbria University, Newcastle, UK
li.zhang@northumbria.ac.uk

Abstract. Real-time contextual affect sensing from open-ended multithreaded dialogue is challenging but essential for the building of effective intelligent user interfaces. In this paper, we focus on context-based affect detection using emotion modeling in personal and social communication context. It focuses on the prediction both of the improvisational mood of each character and emotional implication in direct related improvisational context during the creative improvisation. Evaluation results indicate that the new developments on contextual affect sensing enabled an affect inspired AI agent to outperform its previous version in affect sensing tasks.

Keywords: affect sensing, multithreaded dialogue, improvisational interaction.

1 Introduction

Recognizing complex emotions, value judgments and moods from open-ended text-based multithreaded dialogue and diverse figurative expressions is a challenging but inspiring research topic. In order to explore this line of research, previously we have developed an affect inspired AI agent embedded in an improvisational virtual environment interacting with human users [1, 2]. The AI agent has been equipped with the capabilities of detecting a wide range of affect (25 affective states), including basic and complex emotions, meta-emotions and value judgments. Up to five characters are involved in one session and they could be creative under the improvisation of loose scenarios such as school bullying and Crohn's disease[1]. The conversational AI agent also provides appropriate responses based on the detected affect from users' input in order to stimulate the improvisation.

However, our previous affect detection has been performed solely based on the analysis of individual turn-taking user input. Thus the context information has been ignored. However, since open-ended natural language input could be ambiguous, sometimes contextual profiles are required in order to further justify the affect implied

[1] In the Crohn's disease scenario, the sick leading character, Peter, needs to discuss pros and cons with friends and family about his life changing operation in order to make a decision. Janet (Mum) wants Peter to have the operation. Arnold (Dad) is not able to face the situation. Other characters are Dave (Peter's best friend) and Matthew (Peter's younger brother).

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 12–23, 2012.
© Springer-Verlag Berlin Heidelberg 2012

by the speaking character. Moreover in the Relevance theory, Sperber & Wilson [3] suggested that: "communication aims at maximizing relevance; and speakers presume that their communicative acts are indeed relevant". They also further stated that effective communication is not only based on the coding and encoding of messages but also regarding to the inferences of the communicative intention of the speaker. Therefore in this paper, we discuss contextual affect sensing integrated with emotion modeling of personal and social context to justify the affect conveyed in emotionally ambiguous input.

The paper is arranged as follows. We discuss related work in section 2 and new development on contextual affect modeling & sensing in section 3. We discuss evaluation results and future directions in section 4.

2 Related Work

Much research has been done on creating affective virtual characters in interactive systems. Emotion theories, particularly that of Ortony, Clore and Collins [4] (OCC), have been used widely therein. Prendinger and Ishizuka [5] used the OCC model in part to reason about emotions and to produce believable emotional expressions. Mehdi et al. [6] combined a widely accepted five-factor model of personality, mood and OCC in their approach for the generation of emotional behaviour for a fireman training application. Gratch and Marsella [7] presented an integrated model of appraisal and coping, to reason about emotions and to provide emotional responses, facial expressions, and potential social intelligence for virtual agents. Aylett et al. [8] also focused on the agent development of affective behaviour planning.

Recently textual affect sensing has also drawn researchers' interests. ConceptNet [9] is a toolkit to provide practical textual reasoning for affect sensing for six basic emotions, text summarization and topic extraction. Shaikh et al. [10] provided sentence-level textual affect sensing to recognize evaluations (positive and negative). They adopted a rule-based domain-independent approach, but haven't made attempts to recognize different affective states from open-ended text input. Although Façade [11] included shallow natural language processing for characters' open-ended utterances, the detection of major emotions, rudeness and value judgements is not mentioned. Zhe and Boucouvalas [12] demonstrated an emotion extraction module embedded in an Internet chatting environment. It used a part-of-speech tagger and a syntactic chunker to detect the emotional words and to analyze emotion intensity. The detection focused only on emotional adjectives and first-person emotions, and did not address deep issues such as figurative expression of emotion. There is also work on general linguistic cues useful for affect detection (e.g. Craggs and Wood [13]).

Context-sensitive approaches have also been used to sense affect and emotion. Ptaszynski et al. [14] developed an affect detection component with the integration of a web-mining technique to detect affect from users' input and verify the contextual appropriateness of the detected emotions. However, their system targeted conversations only between an AI agent and one human user in non-role-playing situations, which greatly reduced the complexity of the modeling of the interaction context.

Our work focuses on the following aspects: (1) real-time affect sensing for basic and complex emotions in open-ended improvisational role-play situations; (2) affect interpretation based on context profiles; and (3) affect detection across scenarios.

3 Exploitation in Contextual Affect Sensing

Since our previous affect detection was performed solely based on the analysis of individual turn-taking user input, the context information was ignored. As discussed earlier, the contextual and character profiles may not only influence the current speaker's emotional expression but also help to derive affect embedded in communication context. Therefore contextual affect detection has drawn our research attention. In this section, we discuss cognitive emotion simulation for personal and social context, and our approach developed based on these aspects to interpret affect from emotionally ambiguous input, especially affect justification of the previously detected 'neutral' expressions based on the analysis of individual turn-taking input.

3.1 Improvisational Mood Modeling in Personal Context

Lopez et al. [15] has suggested that context profiles for affect detection included social, environmental and personal contexts. In our study, personal context may be regarded as one's own emotion inclination or improvisational mood in communication context. We believe that one's own emotional states have a chain reaction effect, i.e. the previous emotional status may influence later emotional experience. We make attempts to include such effects into emotion modeling. Bayesian networks are used to simulate such personal causal emotion context. In the Bayesian network example shown in Figure 1, we regard the first, second and third emotions experienced by a particular user respectively as A, B and C. We assume that the second emotion B, is dependent on the first emotion A. Further, we assume that the third emotion C, is dependent on both the first and second emotions A and B. In our application, given two or more most recent emotional states a user experiences, we may predict the most probable emotion this user implies in the current input using a Bayesian network.

Briefly, a Bayesian network employs a probabilistic graphical model to represent causality relationship and conditional (in)dependencies between domain variables. It allows combining prior knowledge about (in)dependencies among variables with observed training data via a directed acyclic graph. It has a set of directed arcs linking pairs of nodes: an arc from a node X to a node Y means that X (parent emotion) has a direct influence on Y (successive child emotion). Such causal modeling between variables reflects the chain effect of emotional experience. It uses the conditional probabilities (e.g. P[B|A], P[C|A,B]) to reflect such influence between prior emotional experiences to successive emotional expression. The following network topology has been used to model personal emotional context in our application.

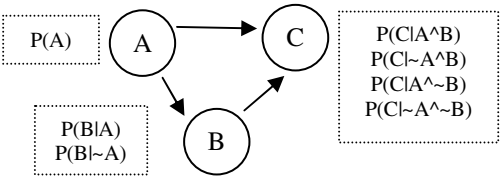


Fig. 1. An emotion Bayesian network

In Figure 1, conditional probabilities are needed to be calculated for the emotional state C given any combination of the emotional states A and B. Theoretically, emotional states A and B could be any combination of potential emotional states. In our application, we mainly consider the following 10 most frequently used emotional states for contextual affect analysis including 'neutral', 'happy', 'approval', 'grateful', 'caring', 'disapproval', 'sad', 'scared', 'threatening', and 'angry'. Any combination of the above emotional states could be used as prior emotional experience of the user thus we have overall 100 (10 * 10) combinations for the two preceding emotions. Also each conditional probability for each potential emotional state given two prior emotional experiences (such as P[happy| A,B], P[approval| A,B] etc) will be calculated. The emotional state with the highest conditional probability is selected as the most probable emotion the user conveys in the current turn-taking. We construct a Bayesian network for each character to sense his/her improvisational mood.

At the training stage, two human judges (not involved in any algorithm development) marked up 3 example transcripts of the Crohn's disease scenario, which consisted of approximately 450 turn-taking inputs. For each character, we extract three sequences of emotions from the improvisation of the 3 example transcripts to produce prior conditional probabilities. We take a frequency approach to determine the conditional probabilities for each Bayesian network. When an affect is annotated for a turn-taking input, we increment a counter for that expressed emotion given the two preceding emotions. For each character, a conditional probability table is produced based on the training data. An example is presented in Table 1.

Table 1. An example conditional probability table for emotions expressed by a particular character

		Probability of the predicted emotional state C being:			
Emotion A	Emotion B	Happy	Approval	...	Angry
Happy	Neutral	P00	P01	...	P09
Neutral	Angry	P10	P11	...	P19
Disapproval	Disapproval	P20	P21	...	P29
Angry	Angry	P30	P31	...	P39

In the above table, the predicted emotional state C could be any of the most frequently used 10 emotions. At the training stage, the frequencies of emotion combinations in a 100 * 10 ((A*B)*C) matrix are produced dynamically. This matrix represents counters (N_{CAB}) for all outcomes of C given all the combinations of A and

B. A one-hundred-element array is also needed to store counters (N_{AB}) for all the combinations of two prior emotions, A and B. Such a conditional probability matrix is constructed at run-time for each human-controlled character in the Crohn's disease scenario based on the training emotional sequences.

For the prediction of an emotion state mostly likely implied in the current input by a particular character at the testing stage, the two prior recent emotional states are used to determine which row to consider in the conditional probability matrix, and select the column with the highest conditional probability as the final output. The emotional sequences used for testing expressed by each character have also been used to further update and enrich the training samples so that these testing emotional states may also help the system to cope with any new emotional inclination because of each character's creative improvisation.

Example conditional probability calculations are shown in the following, where N represents the total number of emotions shown so far by one character and N with a subscript indicates the number of a specific emotion shown given previously expressed emotions. E.g., $N_{happy\_neutral\_happy}$ indicates the occurrences that two prior emotions A and B are respectively happy & neutral and the subsequent emotional state C is happy.

$P(A = happy) = N_{happy}/N$
$P(B = neutral| A = happy) = N_{happy\_neutral}/N$
$P(C = happy| A = happy, B = neutral) = N_{happy\_neutral\_happy}/N_{AB}$

An example algorithm of the Bayesian affect sensing is provided in the following. For the initial run of the algorithm at the testing stage, emotions A, B and C are initialized with the most recent three affects detected for each character purely based on the analysis of individual input.

```
Function Bayesian_Affect_Prediction
{
  1. Verify the contextual appropriateness of the affect
     C predicted by the Bayesian reasoning;
  2. Produce the row index, i, for any given combination
     of the two preceding emotional states A & B in the
     matrix;
  3. Indicate the column index, j, for the recommended
     affect C;
  4. Increment N_AB[i] and N_CAB[i][j];
  5. Update two preceding emotions by:
     Emotion A = Emotion B; Emotion B = The newly
     recommended affect C;
  6. Produce the new row index, k, for any given
     combination of the updated two preceding emotional
     states A & B;
  7. Calculate probabilities (i.e. P[C|A,B] =
     N_CAB[k][column]/N_AB[k]) for the predicted emotional
     state C being any of the 10 emotions;
```

```
  8. Select and return the affect with the highest
     probability as the predicted affect C;
}
```

Pseudo-code for affect prediction using a Bayesian network

In detail, at the testing stage, when an affect is predicted for a user's input using the Bayesian network, the contextual appropriateness of the detected affect will be further justified. The verification processing using neural network-based reasoning (see the following section) results in a final recommended affect. Then the conditional probability table obtained from the training stage is updated with the newly recommended affect and its two preceding emotions. We increment both the counter, N_{CAB}, storing the number of occurrences of the expressed emotion given the two prior implied emotional states and the counter, N_{AB}, for the storage of the frequencies of the two preceding emotions. The next step is to update the two preceding emotions by shifting the previous emotion B to emotion A and replacing emotion B with the newly recommended detected affect C for future prediction. Then conditional probabilities are calculated for the new-round predicted emotion C given the two newly updated preceding emotions. In our application, frequencies are stored from training to testing and are used to form probabilities for prediction when required. The above processing is iterative to predict affect throughout an improvisation for a particular character based on his/her personal emotional profiles.

We extract the following example interaction from the Crohn's disease scenario. Based on the affect detection purely from the analysis of each individual input, we assigned an emotional label for each of the following input as the first step.

1. Arnold: son, let's not think about the operation now. What's for dinner love. [disapproval]
2. Janet: no no peter needs to know what's goin to happen to him darling. [disapproval]
3. Peter: dad, I cannot leav it for later. [disapproval]
4. Dave: what's ur symptoms. [neutral]
5. Janet: peter cannot eat too much. Don't be so rude, dave. [disapproval]
6. Arnold: sweetheart, please I don't want to talk about it now. [disapproval]
7. Janet: How u think peter feels HUH!!! [angry]
8. Matthew: yeah, peter is more important [approval]
9. Janet: Arnold, u just think about urself! [neutral] -> [angry]

We use the sequence of emotions expressed by Janet to illustrate how the contextual affect sensing using Bayesian networks performs. Affect annotation based on the analysis of individual turn-taking input has derived 'disapproval', 'disapproval' and 'angry' respectively for Janet's 2^{nd}, 5^{th} and 7^{th} inputs. 'Neutral' has been detected for Janet's very last input, 9^{th} input. In our application, the context-based affect sensing will run as a back stage processing to monitor the affect interpretation from the analysis of each individual input and it will especially make a prediction when 'neutral' is detected based on the analysis of the users' input itself. As we mentioned earlier, the Bayesian affect sensing algorithm uses the most recent three emotions

experienced by each character to initialize the two preceding emotions A & B and the subsequent emotion C. Thus the two preceding emotions A & B are respectively 'disapproval' and 'disapproval', while the subsequent emotion C is 'angry'. We increment both the counter of the frequencies of $N_{disapproval\_disapproval}$ and the counter of the occurrences of $N_{disapproval\_disapproval\_angry}$. Then the system updates the two preceding emotions by shifting the emotion B to the emotion A and replacing the previous emotion B with the subsequent emotion C, i.e. the two newly updated preceding emotions A & B are respectively 'disapproval' and 'angry'. Then conditional probabilities, P[C| disapproval, angry] are calculated for the predicted emotional state C given the two preceding emotions based on the frequencies obtained from the training data. The emotional state with the highest probability is regarded as the most probable emotion implied in the Janet's next input. The output indicates that 'anger' is implied in Janet's input "Arnold, u just think about urself!". Therefore, the affect annotation changes from 'neutral' to 'angry'.

Since our processing is iterative, the contextual appropriateness of the detected affect 'angry' is further verified by the social context inference using neural networks. The counter of $N_{disapproval\_angry}$ is also incremented. If the inference of the communication context further strengthens the prediction, we also increment the counter for the emotion 'angry' given two preceding emotions as 'disapproval' and 'angry', i.e. $N_{disapproval\_angry\_angry}$. Otherwise, the counter for any recommended other affect supported by the social context inference given the same two preceding emotions is incremented. Thus the testing emotional profiles are used to update the frequencies obtained from the training set to improve the algorithm's generalization abilities for future prediction.

In this way, the AI agent is capable of predicting the improvisational mood of each character throughout the improvisation. We have provided evaluation results for the performances of the improvisational mood modeling using the Bayesian networks with 4 transcripts of the Crohn's disease scenario for the 3 leading characters, Janet, Peter and Arnold in the evaluation section.

However, since the Bayesian inference gathers frequencies of emotional sequences throughout improvisations, it relies heavily on the probability table produced based on such frequencies for future prediction. I.e. it gives high prediction for emotions with high occurrences following any two given preceding emotions and ignores the responding to local emotional environments. For example, the Bayesian network is generally less likely to recommend an emotional state with low frequencies following any two given preceding emotions comparing with other subsequent emotions with high occurrences even though the local interaction context's strong disapproval. Thus it may lead to affect sensing errors.

Moreover, usually, directors take the channels of interventions to guide the drama improvisation in our application. A few story subthemes embedded in the scenario are normally used for interventions to direct the drama improvisation such as, "you could discuss diet and how disease affects it", "ok team, you are in a modern restaurant", "your waiter is Dave today, who has a BIGGGG attitude problem!" etc. Although characters' improvisation is creative and the improvisation for most scenario subthemes contains mixed negative & positive social discussion context,

comparatively for the discussion of a couple of topics, the local interaction context indicates more positive/negative implication than other discussion subthemes. For example, when involved in the discussion of the subtheme, "you are now in a modern restaurant", human actors normally showed excitement for the fancy restaurant & food choices and more positive interactions were contributed. Also when engaged in dialogue context of "don't forget how bowel disease could affect diet, guys", more arguing and accusation situations occurred. The above discussed Bayesian approach seems struggling to reflect such local social emotional implication effectively although it learns about the general emotional inclination in a global manner for each character. In the following section, we discuss a local social affect interpreter based on direct related interaction context using neural network-based inference to provide another justification channel for contextual affect analysis.

3.2 Emotion Prediction in Social Communication Context

As stated earlier, another effective channel is needed to sense affect implication in the most related social interaction context to complement the Bayesian inference. Although directors' interventions sometimes indicate emotional implication of the discussion context to some degree, creative diverse improvisation of actors makes directors employ different topics for interventions in order to stimulate improvisation. Even worse, director's intervention varies from one human director to another. Thus we need a more effective inference channel to sense affect embedded in the communication context. The most recent interaction context contributed by several participant characters has become a useful reliable source of modeling to reveal its general social affect implication.

In our application, the improvisational social interaction context has great potential to contain rich emotional and communicative intentions between speakers and audience. Such emotional communicative intention may also affect the emotional expression of the current speaking character. For example, a recent negative context contributed by the big bully may cause the bullied victim and his/her friends to be 'angry'. Thus we employ a neural network algorithm, Adaptive Resonance Theory (ART-1) to model such social emotional context and sense the positive/negative/neutral implication embedded in such context.

Briefly, ART is a collection of models for unsupervised learning and mainly used to deal with object identification and recognition. ART-1 in particular has the ability to maintain previously learned knowledge ('stability') while still being capable of learning new information ('plasticity'). Such capabilities are especially useful to our application, since they may allow the emotional social context modeling to perform across scenarios. Also, ART-1 has another advanced ability to create a new cluster when required with the assistance of a vigilance parameter. This parameter may help to determine when to cluster an emotion feature vector to a 'close' cluster or when a new cluster is needed to accommodate this emotion vector.

In our application, we use the binary evaluation values (positive and negative) and neutralization implied by the most recent several turn-takings as input to ART-1. In detail, for each user input, we convert its emotional implication into pure

positive/negative and use 0 to represent an absent binary evaluation value and 1 for the presence of a binary evaluation value. For example, for the input from Arnold in the Crohn's disease scenario, "dont boss me about, wife [angry]" when the wife character, Janet, was too pushy towards the husband character, Arnold. We use '0 (neutral), 0 (positive), and 1 (negative)' to indicate the emotional inclination ('angry' -> 'negative') in the user input. In our application, we considered 4 most recent turn-takings as the direct context as inputs to ART-based reasoning. The reasons are as follows. In one session, there are 5 characters involved in the improvisation. If all the 5 characters contribute to the improvisation equally, then except for the current speaker, the 4 most recent inputs are regarded as the updated direct context. Thus, we have taken the 4 recent turn-takings to predict their influence (positive/negative/neutral context) to the current speaker. It is also easily adaptable to accommodate other social context modeling.

In the previous example transcript from the Crohn's disease scenario shown in the above section, for the very last input (the 9th input from Janet), we previously interpreted 'neutral' based on the analysis of the input itself. The personal emotional context prediction based on Bayesian inference derives 'angry' as the most probable affect implied in it. As mentioned earlier, the verification of the contextual appropriateness of the detected affect based on the social context inference is required to further justify the Bayesian prediction.

Thus in the above example interaction, we take the previous four inputs, from Janet (5th & 7th input), Arnold (6th input) and Matthew (8th input), as the direct related social context. Since Janet implies 'disapproval' in the 5th input (binary value combination for neutral, positive and negative: 001) by saying "peter cannot eat too much. Don't be so rude, dave" and shows 'angry' (001) in her 7th input "How u think peter feels HUH!!!", Arnold also indicating 'disapproval' (001) in the 6th input: "sweetheart, please I don't want to talk about it now.", followed by an 'approval' (010) 8th input from Matthew "yeah, peter is more important ", we have used the following emotion vector to represent this most related discussion context: '001 001 001 010 (Janet: 5th, Arnold: 6th, Janet: 7th and Matthew: 8th)'. This emotion feature vector is used as the input to ART-1 to determine if the most related context implies 'positive/negative/neutral'.

Briefly, we begin the algorithm with a set of unclustered emotional context feature vectors (emotional context like the above: '001 001 001 010') and some number of clusters (positive/negative/neutral categories). For each emotional feature vector, it makes attempts to find the cluster to which it's closest. A similarity test and a vigilance test calculate how close each emotional feature vector to the positive/negative/neutral cluster vectors. If an emotional feature vector fails the similarity or vigilance test for all the available clusters, then a new cluster is created for this emotion vector. In our application, we gradually provide emotional context feature vectors to ART-1, which will not only remain the previous classification of positive or negative context in a particular scenario, but also indefinitely integrate new positive/negative context extracted from other interactions across scenarios. For example, if we have the following emotional contexts, from the Crohn's disease scenario, classified previously by the algorithm into three categories, then these categorizations are retained in future predictions.

Class 0 contains: [1 0 0 0 0 1 0 0 1 0 0 1] negative1 (neutral, sad, disapproving and sad)
Class 1 contains: [1 0 0 0 1 0 1 0 0 0 0 1] positive2 (neutral, caring, neutral and disapproval); [1 0 0 1 0 0 1 0 0 1 0 0] neutral1 (neutral, neutral, neutral and neutral)
Class 2 contains: [0 1 0 0 1 0 0 1 0 1 0 0] positive1 (happy, grateful, happy and neutral)

Since ART-1 is not aware which label it should use to mark the above categorization although it classifies the emotional feature vectors based on their similarities and differences and achieves the above classification, a simple algorithm is used to assign labels (positive/negative/neutral context) to the above classification based on the majority vote of the evaluation values of all the emotional states shown in each emotional feature vector in each category. For example, Class 1 has assigned 2 'emotional' vectors and most of the emotional states in all the feature vectors in this category are 'neutral', therefore it is labeled as 'neutral context'. Similarly Class 0 is recognized as 'negative context' with Class 2 identified as 'positive context'. If we add the above example context from the Crohn's disease scenario as a new emotion vector, '001 001 001 010' (contributed by Janet, Arnold, Janet and Matthew), to the algorithm, we have Class 0 updated to accommodate the newly arrived emotional vector as output. Thus this emotion feature vector is 'classified' as a 'negative context'. The 'anger' emotion predicted by the Bayesian inference in the 9^{th} input is verified as the appropriate emotion embedded in a consistent 'negative' interaction context.

4 Evaluations

We carried out user testing with 220 secondary school students from UK schools for the improvisation of school bullying and Crohn's disease scenarios. Generally, our previous statistical results based on the collected questionnaires indicate that the involvement of the AI character has not made any statistically significant difference to users' engagement and enjoyment with the emphasis of users' notice of the AI character's contribution throughout. Briefly, the methodology of the testing is that we had each testing subject have an experience of both scenarios, one including the AI minor character only and the other including the human-controlled minor character only. After the testing sessions, we obtained users' feedback via questionnaires and group debriefings. Improvisational transcripts were automatically recorded to allow further evaluation.

We also produce a new set of results for the evaluation of the updated affect detection component with context-based affect detection based on the analysis of some recorded transcripts of Crohn's disease scenario. Generally two human judges (not engaged in any development stage) marked up the affect of 400 turn-taking user input from the recorded 4 transcripts (not used for the Bayesian training) from this scenario. In order to verify the efficiency of the new developments, we provide Cohen's Kappa inter-agreements for the AI agent's performance with and without the new developments for the detection of the most commonly used 10 affective states.

The agreement for human judge A/B is 0.45. The inter-agreements between human judge A/B and the AI agent with the new developments are respectively 0.43 and 0.35, while the results between human judge A/B and the previous version of the agent are respectively 0.39 and 0.29. Although further work is needed, the new developments on contextual affect sensing improve the agent's performance of affect detection in the test transcripts comparing with the previous version.

We have also provided evaluation results of the improvisational mood modeling using the Bayesian networks for the 3 leading characters in the Crohn's disease scenario in Table 2 based on the analysis of the 4 testing transcripts. We have converted the recognized affective states into binary evaluation values and obtained the following accuracy rates by comparing with the annotation of one human judge.

Table 2. Accuracy rates of improvisational mood modeling using Bayesian networks

	Positive	Neutral	Negative
Janet	63%	29%	94%
Arnold	50%	33%	89%
Peter	46%	36%	72%

Generally negative emotions are well detected across testing subjects. Since in the Crohn's disease scenario, characters are struggling to make decisions about Peter's life changing operation, the improvisation tends to be filled with negative emotional expressions such as worrying, anger and fear. Although positive emotions and neutral expressions are recognized less well, the percentages of the inputs indicating positive and neutral expressions based on the human judges' interpretation are respectively approximate 34% and 25%. Thus although there is room for further improvements, the performances of affect sensing from positive and neutral expressions are acceptable. We also obtain Cohen's Kappa inter-agreement for the indication of the performance of the social context inference using neural networks. In good cases, it achieves 0.71 inter-agreement with one human judge, which is very close to the inter-agreement between the two human judges: 0.91. But there are still some cases: when the human judges both believed that user inputs carried negative/positive affective states (such as angry, disapproval, grateful, etc), the AI agent regarded them as neutral. One most obvious reason is that sometimes Bayesian networks failed to predict some of the positive affective states (e.g. grateful) due to their low frequencies presented in the training data. Also affect sensing based on the analysis of individual turn-taking input sometimes failed to uncover the affect embedded in emotionally ambiguous input due to characters' creative improvisation and diverse channels for emotional expression. We also aim to extend the evaluation of the context-based affect detection using transcripts from other scenarios.

Overall, we have made initial developments of an AI agent with emotion and social intelligence, which employs context profiles for affect interpretation using Bayesian and unsupervised inference. Although the AI agent could be challenged by the rich diverse variations of the language phenomena and other improvisational complex context situations, we believe these areas are very crucial for development of effective intelligent user interfaces and our processing has made promising initial steps towards these areas.

References

1. Zhang, L., Gillies, M., Dhaliwal, K., Gower, A., Robertson, D., Crabtree, B.: E-drama: Facilitating Online Role-play using an AI Actor and Emotionally Expressive Characters. International Journal of Artificial Intelligence in Education 19(1), 5–38 (2009)
2. Zhang, L.: Exploitation on Contextual Affect Sensing and Dynamic Relationship Interpretation. ACM Computer in Entertainment 8(3) (2010)
3. Sperber, D., Wilson, D.: Relevance: Communication and cognition, 2nd edn. Blackwell, Oxford (1995)
4. Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, Cambridge (1998)
5. Prendinger, H., Ishizuka, M.: Simulating affective communication with animated agents. In: The Proceedings of Eighth IFIP TC.13 Conference on Human–Computer Interaction, Tokyo, Japan, pp. 182–189 (2001)
6. Mehdi, E.J., Nico, P., Julie, D., Bernard, P.: Modeling character emotion in an interactive virtual environment. In: Proceedings of AISB 2004 Symposium: Motion, Emotion and Cognition, Leeds, UK (2004)
7. Gratch, J., And Marsella, S.: A domain-independent framework for modeling emotion. Journal of Cognitive Systems Research 5, 269–306 (2004)
8. Aylett, A., Louchart, S., Dias, J., Paiva, A., Vala, M., Woods, S., Hall, L.E.: Unscripted Narrative for Affectively Driven Characters. IEEE Computer Graphics and Applications 26(3), 42–52 (2006)
9. Liu, H., Singh, P.: ConceptNet: A practical commonsense reasoning toolkit. BT Technology Journal 22 (2004)
10. Shaikh, M.A.M., Prendinger, H., Mitsuru, I.: Assessing Sentiment of Text by Semantic Dependency and Contextual Valence Analysis. In: Paiva, A.C.R., Prada, R., Picard, R.W. (eds.) ACII 2007. LNCS, vol. 4738, pp. 191–202. Springer, Heidelberg (2007)
11. Mateas, M.: Interactive Drama, Art and Artificial Intelligence. Ph.D. Thesis. School of Computer Science. Carnegie Mellon University (2002)
12. Zhe, X., Boucouvalas, A.C.: Text-to-Emotion Engine for Real Time Internet Communication. In: Proceedings of International Symposium on Communication Systems, Networks and DSPs, Staffordshire University, UK, pp. 164–168 (2002)
13. Craggs, R., Wood, M.: A Two Dimensional Annotation Scheme for Emotion in Dialogue. In: Proceedings of AAAI Spring Symposium: Exploring Attitude and Affect in Text (2004)
14. Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R., Araki, K.: Towards Context Aware Emotional Intelligence in Machines: Computing Contextual Appropriateness of Affective States. In: Proceeding of IJCAI 2009 (2009)
15. López, J.M., Gil, R., García, R., Cearreta, I., Garay, N.: Towards an Ontology for Describing Emotions. In: Lytras, M.D., Damiani, E., Tennyson, R.D. (eds.) WSKS 2008. LNCS (LNAI), vol. 5288, pp. 96–104. Springer, Heidelberg (2008)

Hybrid Document Matching Method for Page Identification of Digilog Books

Jonghee Park[1] and Woontack Woo[2]

[1] GIST U-VR Lab. Gwangju, 500-712, S. Korea
jpark@gist.ac.kr
[2] KAIST GSCT UVR Lab, Daejeon 305-701, Korea
wwoo@kaist.ac.kr

Abstract. *Digilog Books* are AR (Augmented Reality) books, which provide additional information by visual, haptic, auditory, and olfactory senses. In this paper, we propose an accurate and adaptive feature matching method based on a page layout for the *Digilog Books*. While previous *Digilog Books* attached visual markers or matched natural features extracted from illustrations for page identification, the proposed method divides input images, captured by camera, into text and illustration regions using CRLA (Constrained Run Length Algorithm) according to the page layouts. We apply LLAH (Locally Likely Arrangement Hashing) and *FAST+SURF* (FAST features using SURF descriptor) algorithm to appropriate region in order to get a high matching rate. In addition, it merges matching results from both areas using page layout in order to cover large area. In our experiments, the proposed method showed similar matching performance with LLAH in text documents and *FAST+SURF* in illustrations. Especially, the proposed method showed 15% higher matching rate than LLAH and *FAST+SURF* in the case of documents that contain both text and illustration. We expect that the proposed method would be applicable to identifying various documents for diverse applications such as augmented reality and digital library.

Keywords: Document matching, augmented reality, Digilog Book, page identification.

1 Introduction

Since electronic books (e-book) started appearing in 1990s, they have been received much attention as possible replacements for paper books due to its mobility and relatively low price. However, most readers still more prefer paper books because of the emotional bond between book and reader, i.e., the aesthetic sensation of physical papers. These days, *Digilog Books* are greatly anticipated as the next generation of e-books because it can provide additional information by stimulating visual, auditory, tactile senses in Augmented Reality (AR) environment [4]. Since *Digilog Books* has been emphasized, tracking *Digilog Books* becomes an important research field because it is one of core procedures used within the specialty of AR and computer vision. Usually, tracking *Digilog Books* utilizing a camera, used for image retrieval,

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 24–34, 2012.
© Springer-Verlag Berlin Heidelberg 2012

are comprised of page identification among many pages and tracking from camera image. Early *Digilog Books* usually have applied a marker tracking using ARToolkit [2]. However, there are two disadvantages in the marker tracking to apply it to published books. In order to utilize ARToolkit in tracking *Digilog Books*, artificial markers have to be attached on each page for finding page number and camera pose. In case of the published books, it would be exacting tasks. Secondly, there is a limit on the number of pages because of restriction in the number of possible patterns. Therefore, it would not be applied to the books that have large number of pages. In order to overcome these disadvantages, many researchers have applied NFT (Natural Feature Tracking) methods to *Digilog Books* because the NFT utilizes printed contents such as illustrations and text. Therefore, it can resolve problems of the marker tracking through natural manner.

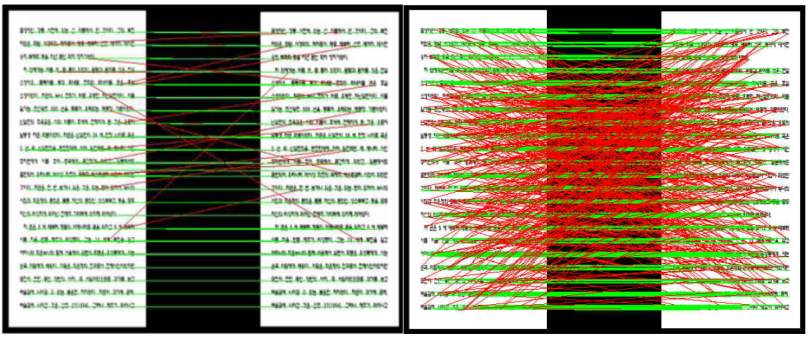
2 Related Works

Generally, there are mainly two categories to identify pages of *Digilog Books* in AR: a marker and a natural feature matching.

First, ARToolkit, which is applied to most of early *Digilog Books*, utilizes marker tracking proposed by Kato and Billinghurst [5]. In order to determine page, the sub-image within a marker is compared by template matching with patterns given the system before. Then, the marker that has the highest matching rate is selected and the designated marker number is the page number. Second, SIFT (Scale Invariant Feature Transform) is a representative method deploying illustrations (texture) in object or scene [7]. However, the processing time for feature extraction of SIFT still takes long time without Graphics Processing Units (GPU). In order to apply NFT to multi-objects tracking in real-time, Wagner applied FAST corners as feature points rather than SIFT features to reduce time in feature extraction [9] [12]. In addition, they trained the images in multi-scale to compensate scale invariant characteristics. Similarly, Kim utilized GPU and multi-threads to reduce feature extraction and matching time [6]. The NFT in Virtual Pop-up Book is another method utilizing template matching [10]. Templates around feature points are used in page determination due to robustness of color histogram in various camera rotations. Another approach in NFT is to utilize text printed on each page. LLAH (Locally Likely Arrangement Hashing) employs each center of words as feature points [8]. Then, geometric relation between features such as cross-ratio is used to generate feature descriptors. Finally, hash indexes that are employed in feature matching are created by the descriptors. Original LLAH has restricted in view point to recognize the page. So, Uchiyama proposed on-line learning method of LLAH to detect pages in various camera viewpoints [11]. However, existing NFT targeting illustrations are not applicable to text because of ambiguity as a result of color similarity. Similarly, NFT targeting text is not applied to illustrations due to unstable repeatability in feature detection. Therefore, if there is a page that consists of only text or illustrations, existing NFT would fail to track the page.

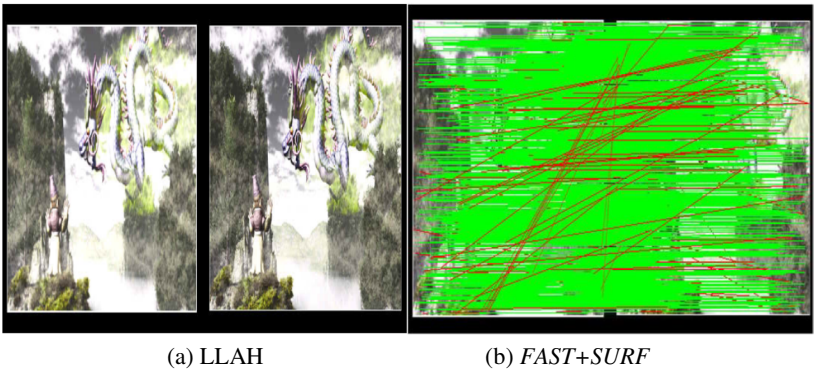
3 Hybrid Document Matching for Page Identification

Generally, NFT methods for illustration are not acceptable to a page which text is printed because of similarity of text descriptors. Figure 1(b) shows a matching result of *FAST+SURF* (FAST features using SURF descriptor) on a text image [1]. In the case, about 1300 features were extracted and only 48% features are correctly matched under 10 degree of rotation. The green lines represent inliers of matching and the red lines mean outliers. However, LLAH classify only few matching correspondences as outliers in figure 1(a). Similarly, NFT methods for text are not appropriate for illustrations. Figure 2(a) shows a matching result of LLAH under 10 degree of rotation. In the case, none of features are matched correctly due to low repeatability of features. In contrast, figure 2(b) shows the matching result of *FAST+SURF* in an illustration.



(a) LLAH (b) *FAST+SURF*

Fig. 1. Matching on a text image with LLAH and *FAST+SURF*



(a) LLAH (b) *FAST+SURF*

Fig. 2. Matching on an illustration image with LLAH and *FAST+SURF*

Figure 3 denotes overall procedure of the proposed method. Firstly, image is divided into two images that one has only text while the other has only illustration. In order to recognize the page, FAST corners and LLAH features are extracted from each image. Then, SURF descriptors are generated from FAST corners and LLAH descriptors are created from LLAH features for matching with pre-trained features.

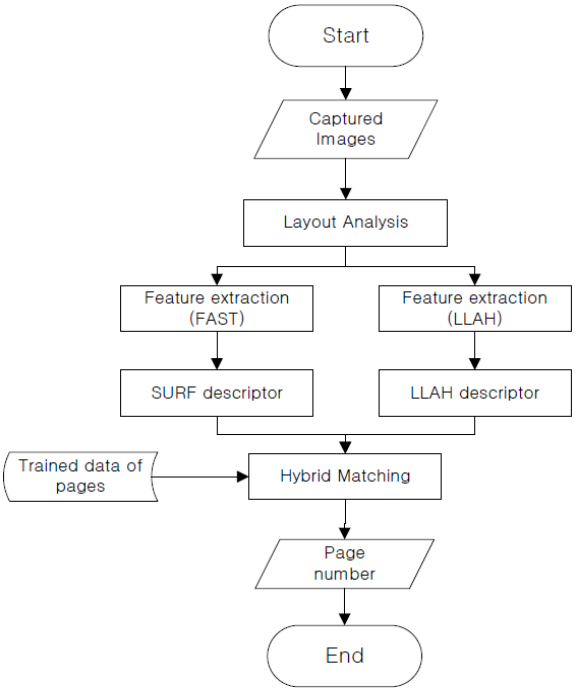


Fig. 3. Flow chart of the proposed method

As the first step, an input image is divided into two parts: illustrations and text by layout analysis. The accurate layout of physical pages is not necessary because the matching is based on feature points. Therefore, rough layout is enough to identify each page of *Digilog Books*. After making a binary image from the image, CRLA (Constraint Run Length Algorithm) is applied to find long white lines in horizontal and vertical direction. From above-mentioned step, common regions in both directions are extracted. In order to make one text line into one block, the white small segments should be removed by applying horizontal direction again [13]. After finding some blocks, we find the contours of blocks except very small blocks that both width and height are less than 5 pixels. After all blocks are detected, each block is categorized into text or illustration. The normalized histogram of each block is utilized to determine the type of the block. Normally, the color of printed text is highly different from background color. Therefore, the histogram has a dominant peak if the background is not mixed up much. If all blocks are categorized, text image and illustration image are generated from an input image according to the block types.

After classifying all blocks, proper features are extracted from each region. FAST corner detector is employed due to a fast processing time. In addition, the reference

images are trained in multi-scale to compensate scale invariant factor. In the text case, we utilize all center points of words like LLAH. First, the text image is transformed into binary image and applies a Gaussian filter to remove noises. Then, dilation is conducted to make words into separated blocks. Finally, the center points of each block are considered as feature points. Figure 4 shows the result of feature extraction of both areas. The blue points denote FAST corners and red points represent LLAH features.

Fig. 4. Feature Extraction Result

After finding matching correspondences from both regions, we need to merge both correspondences together. In the following equation, *I* is a correspondence set from text region and *T* is from illustration. Firstly, each set is sorted to satisfy following condition like PROSAC [3].

$$i_j, i_k \in I : j < k \Rightarrow q(i_j) \geq q(i_k), t_j, t_k \in T : j < k \Rightarrow p(t_j) \geq p(t_k) \tag{1}$$

q and *p* represent correspondence quality function of illustration region and text region respectively. $q(ij)$ is defined as $dist_{2nd}/dist_{1st}$ where $dist_{1st}$ is the nearest vector distance in k-d tree and $dist_{2nd}$ is the second nearest vector distance. For the text region, $p(tj)$ is defined as V/L_{size} where *V* denotes the voting count when the point matched and L_{size} is the list size. After sorting each correspondence set, we create a set M by selecting correspondence alternately like under the following assumption.

$$M = I \cup T, M_{2i-1} = I_i, M_{2i} = T_i, min(\#ofI, \#ofT) > i > 0 \tag{2}$$

$$q(i_j) \geq q(i_k) \Rightarrow P(i_j) \geq P(i_k), p(t_j) \geq p(t_k) \Rightarrow P(t_j) \geq P(t_k) \tag{3}$$

The reason why we select points from different region alternately is that homography from wide area is better than homography from small area. Next, we generate M_n

which is subset of M and the number of elements is n. Then, M_n can be thought as the representative correspondence set which contains highest quality. Finally, initial homography is obtained from M_n by PROSAC. Figure 5 shows matching results on a page consisting of text and illustration by three different methods.

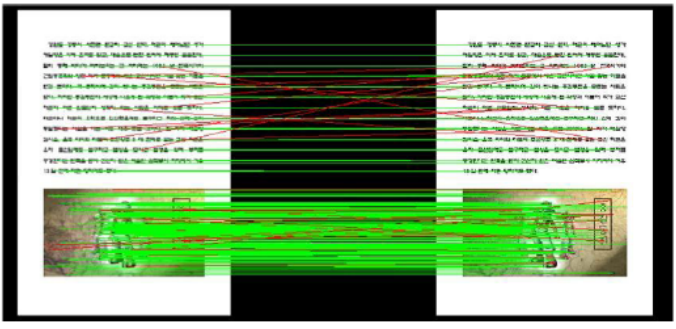
4 Experiments

We experiment the proposed method with LLAH and *FAST+SURF* in various layouts, different type of pages, and various noise levels. Intel Xeon 2.66GHz CPU and 3GB RAM were used for all experiments. Basically, the all image resolution was 640*480 and Gaussian smoothing was utilized for simulating noise test. In *FAST+SURF*, FAST (n=12) corners extracted with non-maximal suppression and 64 feature vectors for each feature was generated by SURF descriptor. In LLAH, 8 neighbor feature points were selected from each feature and each feature vectors discriminated by 15 levels that maximum value was 10. In addition, hash size was 200. The matching rate was defined as *"The number of inliers / The number of matching correspondences"*.

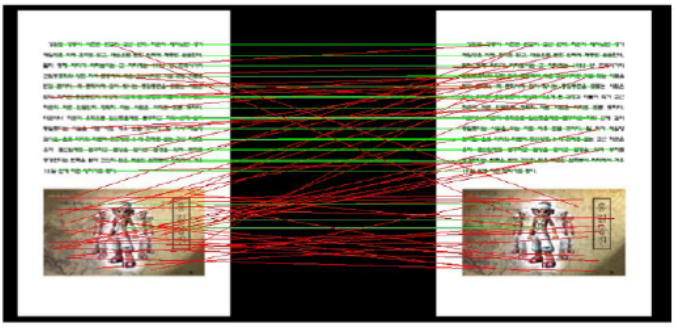
In order to check the performance of matching rate, we tested three methods on three types (text, illustration, both text and illustration) of image set without noise. The 9 test sets consisting of text images (two columns) was tested and each image was rotated by 10 degree increments along x axis. Therefore, 81 images, which content was only text, were tested. All methods recognized successfully from 10° to 50° and LLAH shown the best performance in text images. The performance of the proposed method was almost same as LLAH (Avg. 1.5% difference) as shown in figure 6. The reason of the difference was algorithm difference of finding matching algorithm.

The test set consisting of illustration images (one large figure) was rotated in the same way as text images. 81 images which content was only illustration were tested as well. In a different way with *FAST+SURF* in text images, LLAH was not able to detect reference images even in small rotation. Both *FAST+SURF* and the proposed method were able to detect pages from 10°to 60°. In addition, *FAST+SURF* showed better performance in illustration images than text images as we expected. The proposed method showed the similar performance (Avg. 0.2% difference) as *FAST+SURF* in illustration images. Figure 7 show the proposed method represents stable results according to the content type of pages.

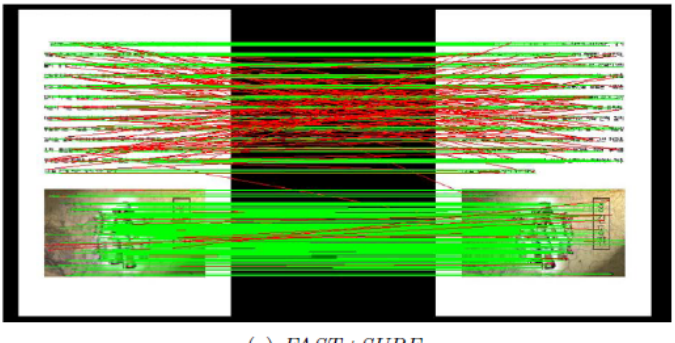
The matching rate of proposed method shown the better performance than *FAST+SURF* (Avg. 16.1% difference) and LLAH (Avg. 19.6% difference) respectively in the case of pages printed both illustration and text. Especially, in the case of 60°, the proposed method was able to detect pages using correspondences from both areas. It denotes that the performance of the proposed method is not just sum of both methods. Figure 7 denotes overall performance of each algorithm according to content type. To sum up, layout based method shows reasonable performance in each type of pages according to the area of the green triangle in the graph. Moreover, the proposed method can be applied to three content types of image stably because its shape is almost regular triangle.



(a) The proposed method.



(b) LLAH.



(c) *FAST+SURF.*

Fig. 5. Comparison of matching result on a same document

The second experiment was conducted in order to check robustness of the proposed method in noise images according to content type of pages. As the first experiment, each method was tested in three types of pages (text, illustration, and text & illustration). Each image was blurred with Gaussian smoothing with difference (from 0.5 to 2.5).

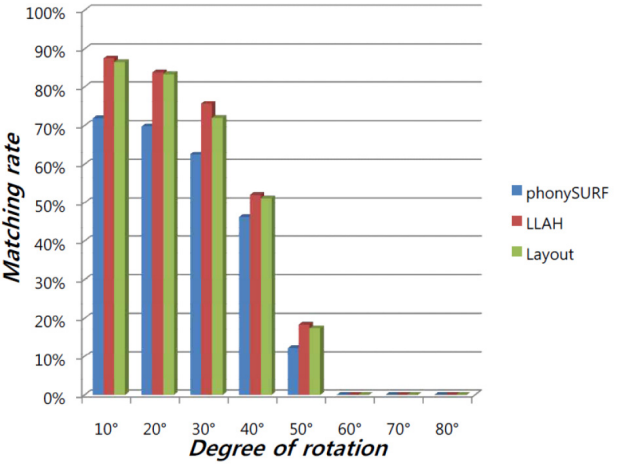


Fig. 6. Comparison of matching result on text images

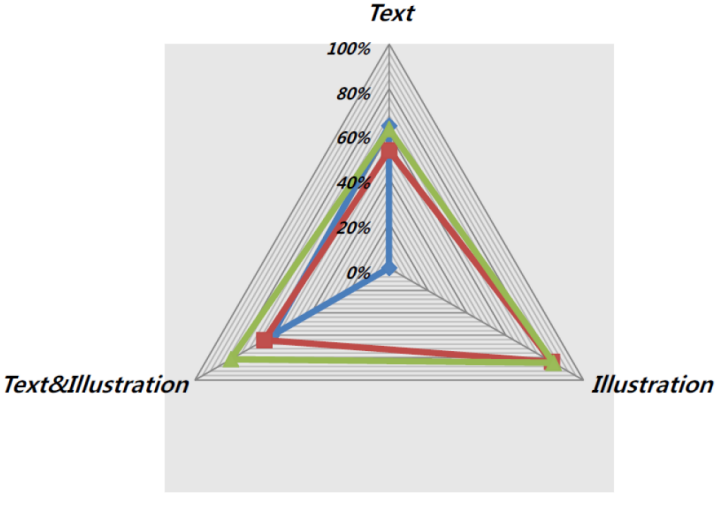


Fig. 7. Matching rate on various content types from 10 to 50 degree of rotation

Figure 8 shows the matching rate of each method with different noise level in text images. Even though *FAST+SURF* showed reasonable performance in the first experiment, this graph verifies why it cannot be applied to text images. The proposed method showed similar result (Avg. 2.6% difference) with LLAH and much better performance than *FAST+SURF* (Avg. 40.8% difference) in text images. The reason why the proposed method has different performance with LLAH was accuracy of layout analysis process. When images are blurred, layout classification did not work correctly. If some blocks were missed in layout analysis process, neighbor features of the blocks also cannot be matched correctly because of high dependency between features. As the result, the matching rate became lower than LLAH.

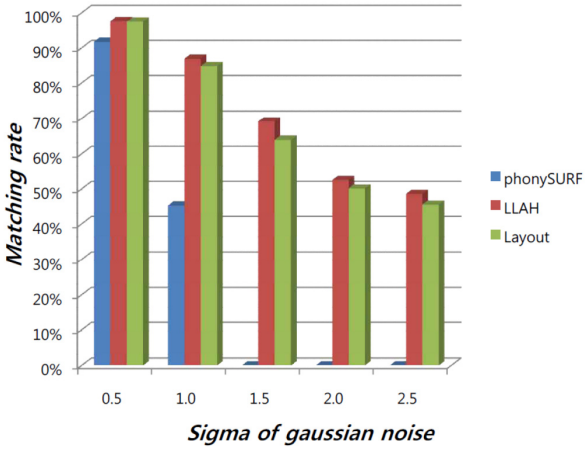


Fig. 8. Matching rate on text images according to various Gaussian noises

Figure 9 denotes matching result under various Gaussian noises on illustration images. As we expected, LLAH was not able to detect the pages like *FAST+SURF* in text. *FAST+SURF* showed quite robust performance under various image noises. The proposed method showed very similar performance from 0.5 to 1.0. However, the gap with *FAST+SURF* was increasing as noise level increased. In the case of large size of illustration, some small blocks were generated inside the illustration because of fixed constraint in CRLA. In addition, when the blocks were classified as a text, the proposed method was not able to match in the blocks. However, the proposed method also shown reasonable performance in various noise level.

In summary of experiment about noise, figure 9 demonstrates the maximum detectable noise level of each algorithm. We decided the detectable noise as 40%

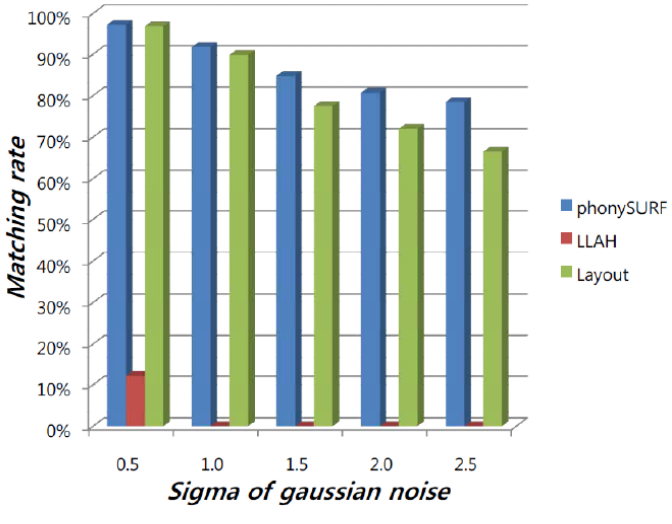


Fig. 9. Matching rate on illustration images according to various Gaussian noises

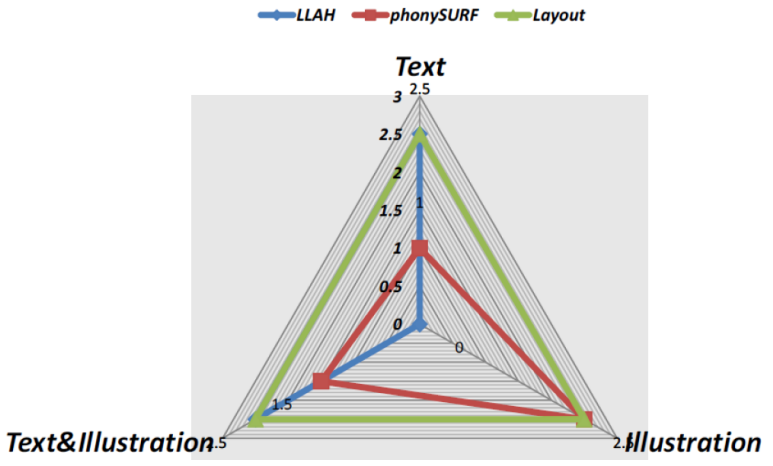


Fig. 10. Maximum detectable noise on various contents

matching rate. As the graph shown, the proposed method showed the most robust matching result in all types of pages. In the case of LLAH, it showed the higher matching rate than *FAST+SURF* except illustration case. The reason why LLAH had much higher matching rate in mixed images was LLAH detected only few correspondences in illustration area. On the other hand, *FAST+SURF* detected many correspondences in text area and most of them were outliers.

5 Conclusion and Future Work

The goal of this research is to transmute existing books into *Digilog Books*. Existing methods such as marker and natural feature, has some limitations to be applied for achieving the goal. Therefore, we proposed layout based matching for *Digilog Books*, which applies proper algorithms according to type of areas (*FAST+SURF* for illustrations and LLAH for text). The proposed method showed higher matching rate if pages consist both text and illustration. Consequently, PROSAC finds homography matrix and removes outliers took less time than single methods in mixed images consisting text and illustration while maintaining accuracy.

Acknowledgments. This research is supported by MCST/Kocca under the CT R&D Program and by NRF/MEST Global Frontier R&D Program on <Human-centered Interaction for Coexistence>(NRF-M1AXA003-2011-0028361).

References

1. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)

2. Billinghurst, M., Kato, H., Poupyrev, I.: The magicbook-moving seamlessly between reality and virtuality. IEEE Computer Graphics and Applications 21(3), 6–8 (2001)
3. Chum, O., Matas, J.: Matching with PROSAC-progressive sample consensus. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 220–226. IEEE (2005)
4. Ha, T., Lee, Y., Woo, W.: Digilog book for temple bell tolling experience based on interactive augmented reality. Virtual Reality, 1–15 (2010)
5. Kato, H., Billinghurst, M.: Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In: Proc. 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR 1999), October 20-21, pp. 85–94 (1999)
6. Kim, K., Lepetit, V., Woo, W.: Scalable real-time planar targets tracking for digilog books. The Visual Computer 26(6), 1145–1154 (2010)
7. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
8. Nakai, T., Kise, K., Iwamura, M.: Use of Affine Invariants in Locally Likely Arrangement Hashing for Camera-Based Document Image Retrieval. In: Bunke, H., Spitz, A.L. (eds.) DAS 2006. LNCS, vol. 3872, pp. 541–552. Springer, Heidelberg (2006)
9. Rosten, E., Drummond, T.W.: Machine Learning for High-Speed Corner Detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
10. Taketa, N., Hayashi, K., Kato, H., Noshida, S.: Virtual Pop-Up Book Based on Augmented Reality. In: Smith, M.J., Salvendy, G. (eds.) HCII 2007, Part II. LNCS, vol. 4558, pp. 475–484. Springer, Heidelberg (2007)
11. Uchiyama, H., Saito, H.: Augmenting Text Document by On-Line Learning of Local Arrangement of Keypoints. In: Proc. 8th IEEE/ACM International Symposium on Mixed and Augmented Reality ISMAR 2009, pp. 95–98 (2009)
12. Wagner, D., Reitmayr, G., Mulloni, A., Drummond, T., Schmalstieg, D.: Pose tracking from natural features on mobile phones. In: Proc. 7th IEEE/ACM International Symposium on Mixed and Augmented Reality ISMAR 2008, September 15-18, pp. 125–134 (2008)
13. Wahl, F., Wong, K., Casey, R.: Block segmentation and text extraction in mixed text/image documents. Computer Graphics and Image Processing 20(4), 375–390 (1982)

Game Design Considerations When Using Non-touch Based Natural User Interface

Mohd Fairuz Shiratuddin and Kok Wai Wong

School of Information Technology, Murdoch University
South Street, Murdoch, Western Australia 6150
{f.shiratuddin,k.wong}@murdoch.edu.au

Abstract. In recent years, the advancement in gaming interface has paved ways for faster and more interactive gameplay. There is now an increasing trend of using Natural User Interface (NUI) for computer games. This has also brought forward new challenges in game design. When designing games using non-touch based NUI, it is not simply replacing existing interaction techniques with the NUI. Many considerations must be addressed to create the effective user experience when using NUI. In this paper, discussions of ten game design considerations have been presented. These ten design considerations can be classified into three main categories. Since the use of NUI is a new phenomenon in games, the design considerations presented in this paper are by no means exhaustive. However, the discussions provided in this paper can reveal potential areas for future research in the field of non-touch based NUI for games.

Keywords: Natural User Interface, game design consideration, user interface, interactive gameplay, non-touch based NUI.

1 Introduction

Game development backdrop is constantly evolving, swayed by the proliferation of gaming hardware interfaces. Current generation of computer games has shown a trend for gaming experiences that produce realistic and "natural" interaction techniques. This trend was started by Nintendo's Wii video gaming console where the Wii Remote (or Wiimote) mimics a player's physical action like swinging a sword, a golf club or a tennis racquet, in which the avatar that the player is controlling is also performing the same action in the game. This was followed by the recent trend and phenomenon, i.e. the introduction of a non- touch based Natural User Interface (NUI) device such as Microsoft Kinect. With a non-touch based NUI device, players are able to interact with/in games using their entire bodies, e.g. making a throwing gesture in a bowling ball game, a lashing gesture in such sword duel game etc. Such non-touch based NUI hardware [1] was able to impact the divergence of the gaming marketplace and at the same time game designers are posed with sub-sequent challenges in new game designs and metaphors. According to Nash [2], "The goal of these new

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 35–45, 2012.
© Springer-Verlag Berlin Heidelberg 2012

interfaces is simple: to reduce the entry barrier for everyone whilst still including traditional gamers", and to accomplish this goal, new game design considerations must be addressed to accommodate non-touch based NUI.

Game design is an integral subset of game development. Adams [3] defines game design as the process of "imagining a game, defining the way it works, describing the elements that make up the game (conceptual, functional, artistic, etc.), and transmitting information about the game to the team who will build it." Brathwaite and Schreiber [4] describe game design as the process of designing the content and rules of a game in the pre-production stage and the design of gameplay, environment, and storyline, characters during production stage. Using non-touch based NUI in games is not simply replacing existing interaction techniques. A lot of new considerations must be carefully thought out, planned and implemented, or otherwise the gameplay will suffer and, could cause frustration and dissatisfaction to the players. This paper discusses non-touch based Natural User Interface (NUI) and the factors that have to be considered by designers of games in designing games that uses this new trend of interface. Since the use of non-touch based NUI is a new phenomenon in games, the factors described in this paper are by no means exhaustive. They will evolve as more and more game developers embraces and utilizes NUI in their games. However, the factors outlined in this paper should give a good start for those considering the use of NUI in their games.

2 Natural User Interface (NUI)

According to Adams [3], the purpose of a User Interface (UI) is that it "creates the player's experience, making the game visible, audible, and playable." The role of UI is of primary importance because from the player's interaction with the UI, he/she perceives whether the game is "satisfying or disappointing, elegant or graceless, fun or frustrating." Adams further stated that since the UI is the intermediary element between the player and the internals of the game, the UI 'knows' all about any supported input and output hardware. The UI translates the player's input and actions in the real world; via Windows, Icons, Menu, Pointing device (WIMP)-based or natural-based, into respective actions in the game world; passing on those actions to the core mechanics, and presents the internal data that the player needs in each situation in visible and audible forms. NUI allows a player to execute relatively natural motions, movements or gestures that they can quickly master and be able to manipulate on-screen content. The objective of NUI is to make the player feels "natural". McMahan et al. [5] consider natural interaction techniques as "…those techniques that mimic real-world interaction by using body movements and actions that are similar to those used for the same task in the physical world."

2.1 Types of Devices for Natural User Interface

Natural interaction can be experienced when interacting with hardware devices that allow player to control the UI directly. The first type consists of devices that allow for

direct manipulation via multi-touch screen of objects in 3D virtual environment (VE). Players can interact in multi-player 3D VEs using single or multi-touch surface hardware. Players interact with the device using their finger/s by touching, poking, prodding, sliding, pinching, spreading, tilting, shaking, or rotate the screen. These features can be found on hardware such as the Apple iPhone, iPod touch, iPad (Figure 1), and etc. The second type is devices that can control the player's motion using remote sensing hardware attached to the player. The devices are either held or attached to the player which translates body motions into game control signals. Common examples include the Nintendo Wiimote and Playstation Move (Figure 2) motion controllers. The third and final device type uses remote sensing features that are able to perform 3D spatial tracking of player's motion and gestures, where players interact in and control the elements of the 3D gaming world with their bodies. Without the player even touching any hardware interface, the NUI system can recognize images, gesture and speech, in various configurations. An example of this kind of device is the Microsoft Kinect (Figure 3) sensor.

Fig. 1. iPad [6]

Fig. 2. Wiimote & Move [7, 8]

Fig. 3. Kinect [9]

3 Game Design Considerations When Using Non-touch Based NUI

Game design starts with an idea and is also often a modification on an existing concept [10, 11, 12]. The game designer usually produces an initial game design document containing the concept, gameplay, feature list, setting and story, target audience, requirements and schedule, staff and budget estimates [10]. Typically, the game design stage produces a documentation called the game design document that will be given to a game development team. The game development team, consisting of programmers, artists, and audio/sound designer who will then turned the game

design document into a working game. During the game development cycle, the work-in-progress game will undergo various iterations and testing. When designing games that are non-touch based NUI oriented, game designers have to make several and specific considerations to ensure players who purchased the game are entertained and satisfied with the gameplay. This paper provides a framework of game design considerations that can be separated into three main categories: user based analysis, function based analysis, and ambient based analysis. In this paper the following considerations have been identified that fall under these three main categories:

1) User based analysis
    - Target Audience
    - Genre
2) Function based analysis
    - Players' Perceived Playability
    - Gaming Space Requirement
    - Suitable Gestures
    - Gestural Control
    - Multi-Player/Single Player
3) Ambient based analysis
    - Speech Control & Audio Elements
    - Iconography
    - Types of Display Screen

All these considerations focus on implications to user experience that is one of the key objectives in game design. The following subsections discuss the game design considerations when using non-touch based NUI.

3.1 Target Audience

Game design is a market driven process. Game designers have to define the kind of experience they want to present and think about the audience who would enjoy that experience. Players purchase a particular genre because they like the type of challenges it offers. Adams [3] categorizes the type of audience for the game market as hard-core versus casual game players, men vs. women, adults vs. children, and girls vs. boys. Adams also discusses the issues that should be well thought-out and implemented to make game more accessible to players with impairments and players of other cultures. Adams further suggested for game designers to question themselves with the following, "Who am I trying to entertain" and once they have the answer to this, the following question follows: "Does this feature entertain a representative player from my target audience?"

When using a non-touch-based NUI for games, game designer must clearly define the target audience. Since current NUI based games tend to target the casual game

players, the game activities cannot be too physical and involve prolonged physical activities because these will quickly tire the players. Game activities should be designed to be in short bursts and yet entertaining, with short breaks in-between. Some examples of non-touch based NUI games are Dance Central, Game Party: In Motion, Your Shape Fitness Evolved, Kinect Sports, Kinect Adventures and Kinectimals.

3.2 Genre

Adams [3] defines genre as "a category of games characterized by a particular set of challenges, independent of the setting or game-world content." The game idea may fall within one or several genres, and designers often experiment with mixing genres [10, 13]. The classic game genres and their related activities and challenges in the games are as summarized in Table 1.

Table 1. Summary of game genres and their related activities and challenges

Genre	Types of activities and challenges
Action games	Physical challenges
Real-time-strategy games	Strategic, tactical, and logistical challenges
Role-playing games	Tactical, logistical, exploration, and economic challenges
Real-world simulations (sports games and vehicle simulations)	Physical and tactical challenges
Construction and management games	Economic and conceptual challenges
Adventure games	Exploration and puzzle-solving challenges
Puzzle games	Logic and conceptual challenges

Referring to Table 1, when designing a game that uses non-touch based NUI, game designers should take into consideration the different types of activities and challenges associated with a specific game genre. Not all types of activities and challenges, and game genres are well suited for NUI as the primary interface. If activities and challenges in a specific game genre are not well designed, players will start losing focus and the gameplay will eventually starts to fall apart and be regarded as tiresome, not entertaining and boring.

For example in the strategy games genre, it is usually presented in a bird-eye-view (Figure 4). Using a point-select-command interaction technique a player usually takes control of constructing buildings, exploring the game area and commanding groups of miniature characters. Given the nature of strategy games being fast-paced, the combination of mouse-keyboard interaction technique is proven to be effective for this game genre. However, if a strategy game or any other types of game genre is to support non-touch based NUI, how would the game design be?

Fig. 3. Screenshots from the Command and Conquer [14], and Company of Heroes [15] RTS game

3.3 Players' Perceived Playability

Despite any game genre, it is key that games are perceived as playable and that players feel that they are making progress when playing long games. According to Sweetser and Wyeth [16], natural interaction techniques affect the player's performance and his overall experience while playing a particular game. An important goal for a computer game is player's enjoyment. Player performance influences his enjoyment. Players will not continue to play game that they do not enjoy. Player performance is thus particularly important because players will experience anxiety if the challenge of a game proves much greater than their own ability to perform [17]. Gee [18] believed that games that people cannot learn to play and from which they do not get the enjoyment of learning would not sell.

Some games were designed to be fairly complex with a steep-learning curve of the controls and were meant for hard-core players. Casual players do not find games with complex control mechanisms to be entertaining and playable. If a game supports NUI, it should be designed in such a way that the control mechanisms feel natural and can almost immediately be mastered without long winded instructions or tutorials. With careful thoughts, planning and design, games that support NUI can be entertaining since NUI promotes the players to be a natural when interacting with games; the player is the controller.

3.4 Gaming Space Requirement

Gaming space requirement is another design consideration. Current non-touch based NUI games have a minimum gaming space requirement with some needing between 6 to 8 feet for optimal gaming experience (Figure 5). In relation to the space requirement, game designers should consider whether the player will be standing, sitting, moving, swinging their hands, stepping their feet etc. As most gaming activities tend to occur at home or in a private confined area, location is also another important factor to consider; whether the play area is in the living room, the bedroom,

or a dedicated game room; and whether these areas can comfortably accommodate the space requirement without too much rearrangement of existing items such as furniture that may be present. The number of players that can effectively occupy the gaming space should also be taken into consideration when designing games that supports non-touch based NUI.

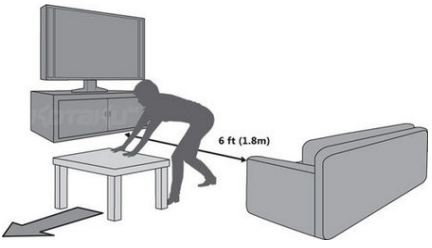


Fig. 4. Gaming space requirement for a typical non-touch based NUI game (Totilo & Crecente, 2010)

3.5 Suitable Gestures

Suitable gestures that will be used when playing non-touch based NUI games should also be taken into account. The gestures performed should be appropriate to whether the game is played in a public setting or more of a private setting. The gestures should also be appropriate for the intended target audience such for adults or for children (Figure 6).

Fig. 5. Examples of gestures for non-touch based NUI games (Makaiookami, 2010; Leavey, 2010)

Nash [2] suggests on limiting the number of gestures that should be performed by the players and can be recognized by the NUI system at any one time. From Nash's user testing experience, he concludes that players can only remember a certain number of gestures and gesture shapes (e.g. circle, oval and zig-zag). To ensure a quick level of adaptability, the gestures and shapes need to be regularly repeated by the player, or at least there must be a non-invasive memory jogger system such as environment shapes or tracing shapes. Nash also posits that "gestures often work best in games where there is a good context for them, e.g. a wizard game such as in the Harry Potter series where you are casting gesture spells".

3.6 Gestural Control

"Gorilla Arm Syndrome" is often associated with touch-screen interface (Carmody, 2010; Boring et al 2009). According to Carmody (2010), the term "Gorilla arm" was coined by engineers circa 30 years ago and that "It's the touch-screen equivalent of carpal-tunnel syndrome." This syndrome was observed mostly with larger touch-screens where movements cause strain on the arm and shoulder after people do tasks with hands up in front of their bodies for long periods of time. Gestural control for non-touch based NUI games should be designed in such ways to avoid the "Gorilla arm" effects since front-end navigation is important and viable for gestural control. Nash [2] believes that "As more titles are released, some control systems are clearly more "usable" by a greater part of the gaming population than others. Development guidance started with handles and rails but seems to be polarizing between the pointers and 'hover' buttons system, and the arm sweep menus and swipe system. Interestingly, approximately 20% of players seem to have difficulty with the best gesture systems indicating that there is still much to learn in this area.

3.7 Multi 3 Player/Singles Player

Non-touch based NUI games should be designed for single player or multiplayer settings. With careful consideration during design, it is also possible to involve not only human-computer interactions but also human-human interactions among the players. Players should be able to naturally collaborate or compete with each other in the same virtual and physical space. The game should also be designed to be a spectator sport. As long as the bystanders stay out of camera view, they can watch the players play. The game should allow players and bystanders to switch places if the current players get tired. This will provide a continuous collaborative experience for a large and dynamic audience.

Laakso and Laakso (2006) developed a non-touch NUI system which they termed as "body-driven" games. Laakso and Laakso provide a design guideline for developing such multi-player games, where certain features should be present in these games. They include:

- Co-operation - Games should encourage players to co-operate in some form, either "team vs. team" or all working towards a common goal. A purely "all vs. all" games tend not to be as addictive in the long run.
- Simplicity - Games should not require the players to position themselves with precise accuracy or perform actions that require very tight timing.
- Intuitiveness - The control metaphors should have reference points to natural human actions.
- Neutrality - Player switches should not be an issue; players should be allowed to enter and leave during an on-going game.
- Scalability - Games should be playable with varying numbers of players, and if possible those numbers should be allowed to change during a game.
- Robustness - Both player approximation and gesture recognition should be simple and robust in order to work under varying conditions.

3.8 Speech Control and Audio Elements

NUI systems should have the ability to use voice commands and controls as one of the means to interact with games. Even though voice recognition technology has evolved significantly throughout the years, the synthesization of continuous natural human speech is far from perfect and cannot be effectively used in games. However, despite this limitation, meaningful single word, phrases and short sentences that have very less pronunciation variations can still be used in games as supporting or alternative method of interaction with games. Nash [2] suggests that game designers make sure that all active words or phrases are very different to one another, and then the system is only trying to identify a limited number of words or phrases. This is to avoid recognition errors because too many words and phrases results in much lower recognition confidence.

Audio elements in a game include sound effects, vibration, ambient sounds, music, dialog and voiceover narration. Adams [3] suggests the inclusion of a facility that allows player to adjust the volume level of music independently from the volume level of other audio effects – including turning one or the other off completely. Some players opt to only hear the sound effects and other sounds, thus turning off the music entirely.

Game designers should include and utilize audio elements and facilities available for a NUI system to ensure players have mostly realistic and authentic experience while playing the game and a memorable experience after playing the game. For example, sound effects correspond to the actions and events of the game world, such as the sound of gunfire or footsteps. While ambient sounds give the player aural feedback, for example, traffic sounds make the player feels that he is in an urban street or sounds of exotic birds inform him that he is in a jungle.

3.9 Iconography

Iconography is the study of identification, description and interpretation of the content of images. Since modern games are highly visually driven, the effectiveness of iconography is of vital importance. This is because the communication of the game, and the subsequent interpretation of the player of that communication, both rely heavily on iconography [2]. For example, icons convey meaningful and identifiable information in a very limited space, thus game designers should make them obvious and unambiguous [3]. Icons should be used for symbolic data that record a small number of possible options. Adams further suggests that icons should be made thematically appropriate that they look as if they belong in the game world. Nash [2] suggests that the effectiveness of the iconographical mechanisms should be tested exhaustively with a usability program. Nash further emphasizes that “the use of one graphic over another can save untold numbers of problems.” To support non-touch NUI based games, icons and buttons should be designed to accommodate effective control of the gameplay. One cannot simply use the standard mouse pointer with NUI. Players may perceived that there is no direct relationship between the mouse pointer

with the gestures they are performing e.g. the action of grabbing and throwing objects in the game world. A cursor icon that looks like a hand that has different hand animations will be more appropriate.

3.10 Types of Display Screen

Display screen for non-touch based NUI has great influence for players to experience maximum satisfaction and enjoyment when playing games. Game designers should always consider the types of display screen (e.g. a HDTV) the players may need to ensure optimal gameplay experience and enjoyment. Since current NUI technologies require players to be at a certain distance and sometime at a certain angle from the TV, NUI games require a large screen sized display no less than 32 inches. If the display screen is too small, the game elements on screen cannot be seen easily hence reducing the level of playability, involvement and immersion.

4 Conclusions

Natural User Interface or NUI has become at the forefront in human-computer interaction in the 21st century, and becoming a significant element of game interface. Non-touch based NUI is quickly becoming the trend of human-game interaction. Since its launch, the NUI Kinect sensor from Microsoft has sold more than 10 million units [1]. This number alone shows that NUI has a great potential not only for games but also for other types of computer application. NUI provides a more intuitive and natural way for players to focus more of game-playing and interaction with the game content. However, introducing a non-touch based NUI into games brings in new challenges for game designers. Using non-touch based NUI in games is not simply replacing existing interaction techniques. Game design considerations as described in this paper must be carefully thought out, planned and implemented, or otherwise the gameplay will suffer and could frustrate the players. As more and more games utilized non-touch based NUI, new considerations when designing games will emerge to counter the challenges that may arise.

References

1. Molina, B.: Microsoft: Kinect sales top 10 million. USA Today (2011), http://content.usatoday.com/communities/gamehunters/post/2011/03/microsoft-kinect-sales-top-10-million/1 (retrieved on March 26, 2011)
2. Nash, J.: The Player Becomes The Producer: Natural User Interface Design Impact (2011), http://www.gamasutra.com/view/feature/6296/the_player_becomes_the_producer_.php?page=1 (retrieved on March 22, 2011)
3. Adams, E.: Fundamentals of Game Design, 2nd edn. New Riders (2010)
4. Brathwaite, B., Schreiber, I.: Challenges for Game Designers. Charles River Media (2009)

5. McMahan, R.P., Alon, A.J.D., Lazem, S., Beaton, R.J., Machaj, D., Schaefer, M., Silva, M.G., Leal, A., Hagan, R., Bowman, D.A.: Evaluating natural interaction techniques in video games. In: 2010 IEEE Symposium on 3D User Interfaces (3DUI), pp. 11–14 (2010)
6. Baros, A.: iPad Full Review + Video (2010), http://techbaros.com/ipad-full-review-video.html (retrieved on April 5, 2011)
7. Pash, A.: Control your computer with a Wiimote (2007), http://lifehacker.com/#!280782/control-your-computer-with-a-wiimote (retrieved on April 5, 2011)
8. Paul, I.: Sony's PlayStation Move: What You Need To Know (2010), http://www.pcworld.com/article/191301/sonys_playstation_move_what_you_need_to_know.html (retrieved on April 5, 2011)
9. Clark, D.: Kinect for Christmas: future unleashed (2011), http://donaldclarkplanb.blogspot.com/2011/01/kinect-for-christmas-future-unleashed.html (retrieved on April 5, 2011)
10. Bates, B.: Game Design, 2nd edn. Thomson Course Technology (2004)
11. Adams, E., Rollings, A.: Andrew Rollings and Ernest Adams on game design. New Riders Publishing (2003)
12. Bethke, E.: Game development and production. Wordware Publishing, Inc. (2003)
13. Oxland, K.: Gameplay and design. Addison Wesley (2004)
14. Ferry: Command and Conquer 3: Control Interface & Side Bar Definitions - Video Games Blogger (2006), http://www.videogamesblogger.com/2006/10/16/command-and-conquer-3-control-interface-side-bar-definitions.htm (retrieved on April 5, 2011)
15. White, D.: Company of Heroes Online Tanks (2011), http://www.thegamereviews.com/article-5304-company-of-heroes-online-does-fatal-nosedive.html (retrieved on April 5, 2011)
16. Sweetser, P., Wyeth, P.: Gameflow: A Model for Evaluating Player Enjoyment in Games. Computers in Entertainment 3(3) (2005)
17. Johnson, D., Wiles, J.: Effective Affective User Interface Design in Games. Ergonomics 46, 1332–1345 (2003)
18. Gee, J.P.: What Video Games Have to Teach us About Learning and Literacy. Palgrave Macmillan (2003)

Effects of the Sequence of Game-Play and Game-Design on Novices' Motivation, Flow, and Performance

Li-Chun Wang and Ming-Puu Chen*

Graduate Institute of Information and Computer Education
National Taiwan Normal University, Taiwan
cct101wang@gmail.com, mpchen@ntnu.edu.tw

Abstract. The purpose of this study was to examine whether providing an active learning approach during game-based learning generates positive impacts on the learners. The effects of learning approaches (active exploration vs. tutorial) on eighth-graders' *Scratch* programming performance, motivation and flow experience were examined in this study. The differences between the active exploration group and the tutorial group included the sequence of game-play and game design activity and the supporting strategy. One hundred and twenty-five junior high school students participated in the study. The results showed that (1) the active exploration approach triggered higher learning motivation and flow experience than the tutorial approach did, (2) the sequence of game-play and game-design did not differentiate learners' project performance, and (3) the causes of why high motivation and flow experience did not bring about better project performance need to be further investigated.

Keywords: game-based learning, motivation, flow experience, programming.

1 Introduction

Game-based learning is an emerging learning approach in today's highly interactive digital world. Digital games play the role of educational contexts to improve learning by motivating the learners and engaging learners in the learning process. Although there has been much written about positive aspects of computer games, it has been confirmed that computer games may cause negative impacts on learning, and therefore, they need to be deliberated designed [1], [2]. Meanwhile, a considerate design of learning approach is the key to enhance learners' engagement in the learning process. Therefore, the instructional designs of game-based learning must contain appropriate learning approaches in order to make learning become more engaging and effective. The barriers novice programmers encountered could be categorized into three aspects: pedagogy, content and technology. The learning activities for programming, learners' learning interests and learning motivation were usually unmatched based on the perspective of pedagogy [3], [4]. Furthermore, novice programmers often face the problem of lack of practice and have difficulty in

* Corresponding author.

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 46–55, 2012.
© Springer-Verlag Berlin Heidelberg 2012

comprehending the syntax and concepts of programming [5]. Researchers suggested that the role of learners should be changed from a game player to a game designer since learners could benefit from the creating process and possess more opportunities to reconstruct acquired knowledge, express ideas and enrich learning experiences [6]. Therefore, the purpose of this study intended to examine whether providing learning approaches, game design activity and game-play activity, generates positive impacts on the learners.

2 Related Literatures

2.1 Game-Based Learning

"Educational games for learning" can be defined as "applications using the traits of video and computer games to create engaging and immersive learning experiences for achieving specified learning goals, outcomes and experiences [7]". Therefore, the purpose of using digital game in educational contexts is to improve learning by motivating the learners [8]. Equipping with the elements of challenge, fantasy, goals, rules, feedbacks, curiosity and control, digital games are motivating to the learners [9], [10], [11]. Previous studies have illustrated positive aspects of computer games as well as negative impacts on learning [1], [2]. For the positive aspect, the effects of game-based learning can be classified into three dimensions including knowledge, skills and affective [10]. From the knowledge-based perspective, games play a role of cognitive tools to (a) provide multiple representations, (b) activate prior knowledge [12], (c) connect game-playing experience and real-life learning experience [13], and (d) provide trial-and-error opportunities for problem-solving and help learners to engage in higher order cognitive thinking [14]. The process of playing games can be regarded as a trial-and-error practice to (a) improve learners' information literacy and skills [7], and (b) provide rich learning experience for problem-solving and transference skills [15]. From the affective perspective, games play a role in motivating learners by adding challenge, feedback, curiosity and fantasy features within the learning process [10], [16]. On the other hand, (a) learners may gain less from learning when game features have been added to subject matters [16]; (b) games may not appeal to every student [12]. These negative aspects imply that game features and specific domain knowledge should be put into careful consideration when designing a game. In this way, the application of games for learning brings promising advantages. Despite the potential advantages mentioned, few studies were conducted to provide empirical evidence for supporting the effectiveness of game-based learning.

Generally speaking, game-based learning, which becomes a promising tool for providing highly motivating learning situations to learners, can be an effective means to assist learners, since learners construct knowledge by playing, maintaining high learning motivation and applying acquired knowledge to real-life problem solving. Game-based learning enables learners to construct knowledge from trial and error with an integration of engaging play, problem solving, situated learning and challenges [14], [17]. Flow experience refers to a situation in which the learners undergo complete absorption or engagement in an activity. A successful game-based

learning was strongly correlated with high degrees of flow experience, as suggested by Kiili [2]. Meanwhile, games can provide rich learning experiences and therefore play an important role in the development of skills, motivation and attitudes [1], [18]. However, whether the sequence of game-based learning activities affects learners' motivation and achievements remains unsolved in previous literature.

2.2 Programming Concepts and Game Design

The ability to reconstruct knowledge, to express ideas creatively and to create information productions can be referred as the capability of information technology, while technological literacy emerged in today's highly interactive digital world can be referred as the programming skills [2]. However, many researchers concluded that novice programmers have difficulties toward programming concepts [5], [19]. The barriers in learning to program could be categorized into three aspects, including pedagogy, content and technology. As Sim [3], [4] suggested, from the pedagogy perspective, programming learning activities were often unmatched with learners' interests and motivation. From the content viewpoint, the abstract concepts of programming were difficult for novice learners. Learners' misconceptions were obstacles for successful conceptual changes, which stemmed from teachers' teachings, learners' experience, intuitions and prior knowledge [20], [21]. Furthermore, for learners to practice programming skills in problem-solving tasks, sufficient technological literacy was required based on the aspect of technology.

For novices to overcome the barriers in learning to program, the role of learners was suggested to be changed from a game player to a game designer [6], [22]. Thus, learners could benefit from the creation process of game design and possess opportunities to actively reconstruct acquired knowledge, express ideas and enrich their learning experiences. Brooks [23] assumed that learning to program involves a mapping between the programming domain and the problem domain, and prior knowledge of programming is the prerequisite to solve problems. Besides, researchers also concluded that pedagogy should be taken into consideration and programming needed to be directed at problem solving instead of rote memorization [21]. By doing so, learners would be able to overcome barriers in learning to program. Therefore, for eliminating the barriers in learning to program, two learning approaches, including tutorial and active exploration, were employed to examine whether learners perform equally in learning programming concepts and their effects on learners' attitudes and motivation.

2.3 Flow Experience and Motivation

Flow experience was defined as "a situation of complete absorption or engagement in an activity" [24]. Csikszentmihalyi [24] also identified eight important factors of enjoyment, including (1) clear goals, (2) immediate feedback, (3) personal skills well suited to given challenges, (4) merge of action and awareness, (5) concentration on the task at hand, (6) a sense of potential control, (7) a loss of self-consciousness, and (8) an altered sense of time. These factors were closely related to learners' internal

experience and external environmental factors. The emerging game-based learning provides an external environment to facilitate learners to concentrate on learning activities and provides internal experience to process higher motivation. Although, strong correlation between learning and flow and positive impact of flow on learning and explorative behaviors were suggested [2], [25], the relationships among learners' behaviors, motivation, and flow experience in game-play need to be further investigated.

3 Research Methods

3.1 Research Design

This study employed a quasi-experimental design to examine the effects of the active learning approach on learners' motivation, flow experience and performance. Multivariate Analysis of Variance (MANOVA) was conducted on learners' performance with a significance level of .05 in the present study. The experiment was conducted in a 7-week session of learning basic programming concepts. The employed learning approaches included the active exploration strategy and the tutorial strategy. The learning activity includes two phases, the game-play phase and the game design phase. As shown in Table 1, the differences between the active exploration group and the tutorial group were the sequence of game-play, game-design activity and the supporting strategy. The active exploration group played the dice game in the first phase. As shown in Fig. 1, the game-play activity served as a learning context for learners to learn programming concepts by means of manipulating game objects through changing corresponding codes. After that, the learners created their own games by designing sprites and actions using *Scratch* scripts, as shown in Fig. 2. The game design phase serves as an application context for learners to apply acquired programming skills in their games. In contrast, the tutorial group received the game design activity in the first phase to create their own games by designing sprites and actions following a step-by-step guidance. Then, in the second phase, game-play of the dice game was employed for learners to observe and relate the actions of game objects and their corresponding codes to enhance learners' programming knowledge and concepts.

Table 1. The design of learning approaches

Learning Approach	Learning activities
Active Exploration	Phase1: Game play Manipulating game objects with corresponding codes Phase2: Game design Designing sprites and actions using *Scratch* scripts
Tutorial	Phase1: Game design Designing sprites and actions using *Scratch* scripts Phase 2: Game play Observing actions of game objects and relate corresponding Codes

```
on (release) {
    left.gotoAndStop(Number(random(6)+1));
    left._rotation = random(360);
    right.gotoAndStop(Number(random(6)+1));
    right._rotation = random(360);
    mid.gotoAndStop(Number(random(6)+1));
    mid._rotation = random(360);
```

Fig. 1. Game-play served as a context for learning and enhancing programming concepts

Fig. 2. Design *a Scratch Game* served as an application context for learners to practice their acquired knowledge and concepts

As shown in Fig. 3, the framework of research design includes pedagogy, content and technology aspects. The aspects are described as follows.

Pedagogy. For the pedagogy aspect, types of learning approaches, including the active exploration approach and the tutorial approach, were employed to engage the learner in the game-based learning activity.

Content. For the content aspect, the game-play activity was conducted to enrich the learner in understanding programming concepts by means of linking actions of game object and the correspondent codes. The content includes the programming concepts, variables and control structure.

Technology. For the technology aspect, the game design activity was implemented using *Scratch* as a tool to empower the learner in trial-and-error practice.

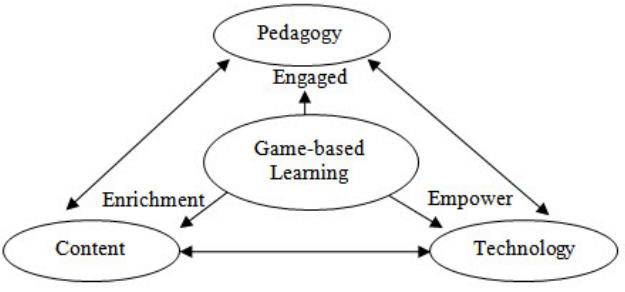


Fig. 3. The research design framework includes pedagogy, content and technology aspects

3.2 Participants

One hundred and twenty-five 8-graders participated in the experiment. All participants were novices to programming languages. Participants were randomly assigned to either the active exploration group or the tutorial group. A programming project, *Design a Scratch Game*, was employed to facilitate participants to apply acquired knowledge to solve real-life problems after learning from the game-play activities. Project performance was assessed for further analysis.

3.3 Instruments

The instruments utilized in the present study included motivation questionnaire, flow experience questionnaire, and project grading rubrics.

The motivation questionnaire was conducted to evaluate learners' intrinsic and extrinsic goal orientation during the game design. The questionnaire was developed by Pintrich, Smith, Garcia, and McKeachie [26] and consisted of eight items on the motivational aspect. Learners were asked to rate themselves on a 5-point Likert-type scale with response options ranged from 1 (strongly disagree) to 5 (strongly agree). The higher scores represented the higher motivation learners possessed. The overall reliability coefficient of the motivation questionnaire was .85 (Cronbach's alpha).

Flow experience questionnaire was conducted to investigate learners' flow experience during game play. The questionnaire consisted of six items. The questionnaire was revised from Kiili and Lainema [25], the flow experience questionnaire and consisted of two subscales, concentration and time disorientation. Learners were asked to rate themselves on a 5-point Likert-type scale with response options ranged from 1 (strongly disagree) to 5 (strongly agree). The higher scores represented the higher level of flow experience that learners experienced and the more concentrated on the game-play. The overall reliability coefficient of the game flow experience questionnaire was .92 (Cronbach's alpha).

The project grading rubrics were employed to evaluate participants' performance in *Scratch* game design activity. The project performance was scored from the correction numbers of control, variables, and control structure after game design activity. The score ranges from 0 (unfinished) to 2 (creativity) of each grading rubric. The higher scores represented the deeper understanding learner achieved and the better fluency learners performed in a game design activity. The overall reliability coefficient of the motivation questionnaire was .87 (Cronbach's alpha).

4 Findings

4.1 Analysis of Learning Approach on Motivation, Flow and Performance

The group means of participants' motivation, flow and performance are shown in Table 2. For learning approach, the active exploration group obtained higher motivation, flow and performance than the tutorial group. The differences between groups were further analyzed as follows.

Table 2. Group means of participants' motivation, flow and performance

Dependent Variable	Group	Mean	SD	N
Motivation	Tutorial	2.990	.855	58
	Active Exploration	**3.424**	.789	67
Flow	Tutorial	3.098	.751	58
	Active Exploration	**3.485**	.776	67
Performance	Tutorial	51.535	16.492	58
	Active Exploration	**55.821**	15.677	67

One-way MANOVA was conducted to examine the effects of learning approach on participants' motivation, flow and performance. First, as shown in Table 3, Levene's tests of equality were not significant for all dependent measures. The null hypothesis that the error variance of the dependent variable is equal across groups was sustained. The MANOVA summary is shown in Table 4, the main effect of type of learning approach is significant on motivation and flow (motivation: $F_{(1,123)}$=8.718, p=.004; flow: $F_{(1,123)}$=7.977, p=.006), while not significant on performance. That is to say, the active exploration group revealed higher motivation and flow experience than the tutorial group. The result is congruent with Kiili [2] and Wang and Chen [27] that

successful game-based learning was strongly correlated with high degrees of flow experience. The results are also consistent with the research hypothesis that the learners perform equally on programming concepts in the game-based learning context. The possible cause may result from that the game-play activity served as a trigger for arousing learners' motivation and trial-and-error practice and engaged the learners during game-play. However, the high motivation and flow experience during game-play were not carried to the game-design stage and generated better project performance for the active exploration learners. Therefore, learners performed equally in the game-design project.

Table 3. Summary of Levene's tests for motivation, flow and performance

Dependent Measure	F	df1	df2	Sig.
Motivation	.331	1	123	.566
Flow	.232	1	123	.631
Performance	.003	1	123	.956

Table 4. MANOVA Summary for group on motivation, flow and performance

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Group	Motivation [a]	5.863	1	4.665	8.718	**.004**	.066
	Flow [b]	4.665	1	5.863	7.977	**.006**	.061
	Performance [c]	571.190	1	571.190	2.215	.139	.018
Error	Motivation	71.931	123	.673			
	Flow	82.726	123	.585			
	Performance	31722.282	123	257.905			

a R Squared = .066 (Adjusted R Squared = .059)
b R Squared = .061 (Adjusted R Squared = .053)
c R Squared = .018 (Adjusted R Squared = .010)

5 Conclusions

Although game-based learning has become a promising strategy for providing highly motivating learning to the learners, whether the sequencing of game-based learning activities affects learners' motivation and achievement remained inconclusive. The present study examined the effects of type of learning approach on learners' motivation, flow experience and performance in game-based learning activities. The findings of this study suggested that the active exploration approach could trigger learners' high motivation and flow experience better than the tutorial approach. As suggested by Kiili [2], Gros [18], Inal and Cagiltay [1], successful game-based learning was strongly correlated with high degrees of flow experience and played an important role in the development of skills, motivation and attitudes. The present study further confirmed that employing game-play activity as an active exploration

approach could bring about high motivation and flow experience in the learners. Therefore, the successful development of skills and attitudes in the learners can be further expected.

As for learning achievement, the active exploration approach and the tutorial approach revealed similar effects on participants' project performance. Both the active exploration approach and the tutorial approach are effective in achieving the same levels of performance. That is to say, the combination of game-play and game-design activities could engage the learner and brought about satisfied learning performance. However, as indicated by Wang and Chen [26], whether the higher degree of flow and motivation facilitate higher achievement remains inconclusive. In the present study, the superior empower effect of the active exploration game-play in enhancing participants' project performance was not found. The PCT (pedagogy, content and technology) design of the learning activities may have balanced the effects of the sequencing effect of the learning activities or the transformation of high motivation in the game-play stage to the game-design stage was not as successful as expected. Therefore, there remains a gap that needs to be further investigated to bridge the high motivation and flow experience with high performance or explain the causes of the failure.

Acknowledgements. This study was sponsored by the National Science Council, Taiwan, under the Grants NSC 98-2511-S-003-034-MY3 and NSC 99-2511-S-003-027-MY3.

References

1. Inal, Y., Cagiltay, K.: Flow Experiences of Children in an Interactive Social Game Environment. British Journal of Educational Technology 38(3), 454–464 (2007)
2. Kiili, K.: Content Creation Challenges and Flow Experience in Educational Games: The IT-Emperor Case. Internet and Higher Education 8, 183–198 (2005)
3. Sim, R.: Interactivity: A Forgotten Art. Computers in Human Behavior 13(2), 157–180 (1997)
4. Sims, R.: An Interactive Conundrum: Constructs of Interactivity and Learning Theory. Australian Journal of Educational Technology 16(1), 45–57 (2000)
5. Norman, D.: The Psychology of Everyday Things. Basic Books, New York (1988)
6. Song, H.D., Yonkers, V.: The Development and Evaluation of Multi-level Assessment Framework for Analyzing Online Interaction (ERIC Document Reproduction Service No. ED442873) (2004)
7. de Freitas, S., Oliver, M.: How Can Exploratory Learning with Games and Simulations within the Curriculum Be Most Effectively Evaluated? Computers and Education 46, 249–264 (2006)
8. Prensky, M.: Digital Game-based Learning. McGraw-Hill, New York (2007)
9. Asgari, M., Kaufman, D.: Relationships among Computer Games, Fantasy, and Learning. In: Proceedings of Educating Imaginative Minds: 2nd Annual Conference on Imagination and Education, Vancouver, BC, Canada, pp. 1–8 (2004)
10. Garris, R., Ahlers, R., Driskell, J.E.: Games, Motivation, and Learning: A Research and Practice model. Simulation & Gaming 33(4), 441–467 (2002)

11. Malone, T.W.: Toward a Theory of Intrinsically Motivating Instruction. Cognitive Science: A Multidisciplinary Journal 5(4), 333–369 (1981)
12. Squire, K.: Video Games in Education. International Journal of Intelligent Games & Simulation 2(1), 49–62 (2003)
13. Pivec, M., Kearney, P.: Games for Learning and Learning from Games. Informatica 31, 419–423 (2007)
14. Adcock, A.: Making Digital Game-based Learning Working: An Instructional Designer's Perspective. Library Media Connection 26(5), 56–57 (2008)
15. Connolly, T.M., Stansfield, M., Hainey, T.: An Application of Games-based Learning within Software Engineering. British Journal of Educational Technology 38(3), 416–428 (2007)
16. Lepper, M.R., Malone, T.W.: Intrinsic Motivation and Instructional Effectiveness in Computer-based Education. In: Snow, R.E., Farr, M.J. (eds.) Aptitude, Learning, and Instruction, III: Conative and Affective Process Analysis, pp. 255–286. Lawrence Erlbaum Associates, Hillsdale (1987)
17. Van Eck, R.: Six Ideas in Search of a Discipline. In: Shelton, B.E., Wiley, D.A. (eds.) The Design and Use of Simulation Computer Games in Education, pp. 31–56. Sense Publishers, Rotterdam (2007)
18. Gros, B.: Digital Games in Education: The Design of Game-based Learning Environment. Journal of Research on Technology in Education 40(1), 23–38 (2007)
19. Ebrahimi, A., Schweikert, C.: Empirical Study of Novice Programming with Plans and Objects. SIGCSE Bulletin 38(4), 52–54 (2006)
20. Mayer, R.E.: Teaching for Transfer of Problem-solving Skills to Computer Programming. In: Corte, E.D., Linn, M.C., Mandl, H., Verschaffel, L. (eds.) Learning Environment and Problem Solving, pp. 193–206. Springer, New York (1992)
21. West, M., Ross, S.: Retaining Females in Computer Science: A New Look at a Persistent Problem. JCSC 17(5), 1–7 (2002)
22. Kafai, Y.B., Ching, C.C., Marshall, S.: Children as Designers of Educational Multimedia Software. Computers and Education 29(2/3), 117–126 (1997)
23. Brooks, R.: Towards a Theory of the Comprehension of Computer Programs. International Journal of Man-Machine Studies 18, 543–554 (1983)
24. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper and Row, New York (1990)
25. Kiili, K., Lainema, T.: Evaluations of an Experiential Gaming Model: The Real Game Case. In: Proceedings of the ED-MEDIA 2006-World Conference on Educational Multimedia, Hypermedia & Telecommunications, pp. 2343–2350. Association for the Advancement of Computing in Education (AACE), Orlando (2006)
26. Pintrich, P.R., Smith, D.A., Garcia, T., McKeachie, W.J.: A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ). MI: National Center for Research to Improve Postsecondary Teaching and Learning (ERIC Document Reproduction Service No. ED338122) (1991)
27. Wang, L.C., Chen, M.P.: The Effects of Type of Game Strategy and Preference-matching on Flow Experience and Performance in Learning to Program from Game-based Learning. Innovations in Education and Teaching International (IETI) 47(1), 39–52 (2010)

Behavioral Traits of the Online Parent-Child Game Players: A Case Study and Its Inspirations

Sujing Zhang[1] and Feng Li[2]

[1] College of Education Science, Hangzhou Normal University, Hangzhou, China
[2] College of Education, Zhejiang Normal University, Jinhua, China
sjzhang@china.com, lifeng_nb@163.com

Abstract. By examining the online parent-child game entitled *Island Survival* , this paper probes into the behavioral traits of parent-child players by means of case study, covering the accumulation and analysis of the dynamic database of these players, and classifies the parent- kid family into three different types: harmonious type, constructive type and crisis type according to the players' behavioral traits, and the way of classifying offers important enlightenment to the design of online parent- child games.

Keywords: Parent-Child Games, Parent-Child Cooperation, Player's Behavioral Traits, Data Analysis.

1 Introduction

The Project of Teenagers' Healthy Personality conducted in 2010 reveals that 75% of high school students interviewed feel they have difficulty in communicating with their parents, or at least occasionally have communication deficiency with their parents, 55.5% of them are not good at associating with other people except their parents; the junior high school students who are unwilling to communicate with parents quarrel with their parents more and more frequently, claiming that their parents can't understand them well. Furthermore, they are afraid of their parents' nagging criticism. [1].All of these phenomena mentioned above can be ascribed to the following two causes. First, teenagers are in the process of "psychological weanling" [2], which implies that the physical, psychological and social development may easily spark conflicts between parents and teenagers, and problems concerning teenagers' academic studies, behaviors and psychology can lead to a vulnerable and sensitive relationship between them [3]; second, parents are still accustomed to the former teaching method implemented in the childhood [4]. As a consequence, they cannot employ more effective approaches to communicate with their children, failing to learn key methods of controlling one's emotions and to renew their notions of parent-child education.

In the past years, our research team, have devoted great efforts to developing parent-child games and creating an interactive environment in which parents and children can

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 56–68, 2012.
© Springer-Verlag Berlin Heidelberg 2012

be of psychological independence and equal communication by observing the rules in games, and this environment is endowed with three functions, i.e., achieving emotional support, offering instrumental assistance (including consultation and suggestion) [5], and obtaining social expectations, i.e., offering certain guidance about which are proper behaviors and which are not.

Island Survival, one of our Online Parent-Child Games, requires both parents and teenagers' participation, and can be played by them in a cooperative way on the same or different computers [6]. Island Survival, in the form of online games, consists of three sections, including *Experience of Parent-Child Education, Parent-Child Collaboration Experience and Visit Exchange among Different Families.*

In the first section, the parent and child will finish branch tasks and become stronger separately and respectively in the parent education experience section and teenager experience section. And the parent will directly experience the perception and disposal modes in family roles as well as positive and negative impacts under such interactive modes, and learn effective parent-child communication skills. By vivid role immersion, the teenager will understand the social anticipations, learn how to control themselves, enhance their sense of responsibility and learn innovation skills.

In the second section, the parent and child will finish the main line task; learn innovation and creation skills by mutual collaboration of the parent and the child in walking through the stage on a mystery isolated island. In addition, different family groups can compete with each other, for there are ranking lists showing how well parents and children cooperate with each other in finishing their tasks. The lists can encourage both parent and child to be better at cooperation and make their relationship more reasonable. This section is also applicable to the single players. However, compared with the cooperative pattern, the single player is unable to obtain doubled rewards and medals, nor is the player able to unlock more missions under the specially-cooperative pattern [7].

As to the third section, Visit Exchange among Different Families, it aims at developing companionship between teenagers and sharing experience about how to nurture children among parents. In this game, parents and children can play together in front of the computer, independently and equally. By cooperating, parents and children can fulfill their tasks, enhance their intimacy between each other, and foster their innovative and creative abilities. More importantly, both the parent and child can retrieve the harmonious relationship and happiness.

In order to find out the players' behavioral traits in the Parent-Child Online Games, we conducted an experimental research in junior high school. Combining both the quantitative and qualitative methods, we accumulate and analyze the dynamic data produced from the players' behaviors. The qualitative research employs the Case Study Method [8], which primarily aims at finding out the players' behavioral traits in the game, and then draws a conclusion of the different family types, which feasibly contributes to a better design of games improving the parent-child relationship.

2 Research Plan

2.1 Research Problems to Be Settled

Our research mainly focuses on the following three points: (1).What kind of behaviors do the players conduct in the Parent-Child Online Game? (2). How many types of family relationship can be concluded on the basis of analyzing the players' behavioral traits? (3). With different types of family relationship as samples, check whether those types of family relationship are in accordance with the degrees of family intimacy based on the adaptive scales(FACESII-CV) or not.

2.2 Research Method

The reason for employing case study is that the players' behaviors in the context of online environments seems to be rather complicated, and the research of the Online Parent-Child Game is just at its beginning stage, lacking substantial and authoritative theories as reference. Therefore, we are not sure enough about the issues possibly occurring in the research. The method of case study has the following features, i.e., it focuses on a deeper understanding [9] and detailed description of subjects [10], and attaches great importance to the discoveries rather than verifications [11].

2.3 Research Sample

In the experiment, we choose students of two classes in a junior high school, totaling 60 pairs of parent-child, as our samples, who are invited to experience a kind of parent-child game for 8 weeks. Afterwards, we collect and analyze the dynamic data from the players' behaviors, and choose four typical pairs of parent-child players, including two girl students and two boy students, for our case study.

2.4 Research Process

Before our experiment, a teacher who teaches information technology course introduces this kind of online game to the students of the two selected classes by telling them the website of this game and giving out guide books for these new players. Then parents and children experience the game at their will in an open environment for 8 weeks, approximately 56 days. They have to play the game around 30 minutes from Monday to Friday, and around two hours at weekends. After the experiment, we test these four families' degrees of intimacy according to the adaptive scales (FACESII-CV), and interview every family for 30-60 minutes.

2.5 Data Collection and Analysis

By obtaining complete materials according to the requirements of the research, an in-depth and comprehensive research can be conducted. Online games integrate the players' mentality and their behaviors. In order to summarize certain regulations from the complicated behaviors of the 60 parent-child samples and to get a commonly-accessible approach, the valuable dynamic data is needed to be accumulated.

The dynamic player data can be categorized into two, one is the storable and renewable database, and another is the temporarily deposited database in the server in the operation of games. These data stored in the database have to be backup by means of snap-shooting, and their corresponding databases at a given time also have to be backup. It is expected that all the snapshots be accumulated within 8 weeks in order to make meaningful statistics and charts accordingly. As to such transient data in the server memory as the data of temporal family teams, we employ a passive information-driven accumulation and an active rule-driven accumulation. The UCINET software is employed as a supplementary tool for analyzing, while the social network analytic method for quantitative research is primarily adopted by means of SPSS software, probing into the difference between the individual cases and overall samples.

3 Research Results

By analyzing snapshots of the database, we get every piece of record of the 60 pairs of parents and kids from the players' server database. Then we choose the roles of ID, degrees of intimacy, index of intimacy, titles of families, time spanning of parent-child's online-state, time spanning of parent-child cooperation, and the period they are in. Altogether we collect seven kinds of data to form the Table1 as follow:

Table 1. A Statistical Table of Parent-Child Players' Behavioral Data

(Level 0=100, Level 1=200, rising to a higher level means the index of intimacy will increase 100 points)

Role ID	Intimacy Level	Intimacy Index	Family Title	Parent-Child Online Time	Parent-Child CollaboratingTime	Age
Lu Haiyang LuHaiyang_p	25	2120	Harmonious Sample and Model Family	52 hours 52 hours	33.2 hours	Robert Age
Wu Haitao Wu Haitao_p	24	1734	Harmonious Model Family	52 hours 51.6 hours	34.7 hours	Robert Age
Jin Hao Jin Hao_p	24	1167	Model Family	51.8 hours 51.1 hours	33.5 hours	Robert Age
…	…	…	…	…	…	…
Zhang Lei Zhang Lei_p	19	1059	Civilized Family	53.2 hours 59.4 hours	25.9 hours	"e" Age
…	…	…	…	…	…	…
Tong Han Tong Han_p	15	566	Learning Family	56.7 hours 67.3 hours	22.8 hours	Machine Age
…	…	…	…	…	…	…
Yi Jinlang Yi Jinlang_p	10	873	Constructive Family	52.2 hours 61.4 hours	16.4 hours	Middle Age
…	…	…	…	…	…	…
Zhen Yun Zhen Yun_p	4	134	Elementary Family	36.7 hours 14.6 hours	2.3 hours	Enlightenment Age
Fang Lei Fang Lei_p	3	322	Elementary Family	37.7 hours 34.2 hours	1.1 hours	Enlightenment Age

From the table above, it can be found that a family with the role ID of Lu Haiyang performs best, because they have a higher degree of parent-child activity, Level 25of intimacy index, and five parent-child medals. Besides, they have entered the most advanced Robert Age after a short time of perfect parent-child cooperation. While the Tong Han family perform poorly, with the highest degree of parent-child activity, Level 15 of intimacy index, three parent-child medals, and the longest time spanning of their online-state, however, their time spanning of parent-child cooperation is only 22.8 hours, and they just narrowly enter the Machine Age; And both Zhen Yun and Fang Lei families' degrees of activities are rather low, with the lowest indexes of intimacy, no parent-child medals. And their time spans of parent-child cooperation are 2.3 hours and 1.1 hours respectively, in the primary Enlightenment Age still.

Based on the analysis of those dynamic data collected from the team cooperation and rivalry between different families, the suggestions and discussions of hot concerns, we can use the UCINNET software to draw a social network analytical graph (as illustrated by Figure 1), which demonstrates the social relationships of the 60 pairs of families taking part in the online game of Island Survival. Every point represents one family, and the lines connecting different points symbolize that these families have once formed a team, cooperated or competed with each other, and the thickness of the lines stands for the frequency of their cooperation.

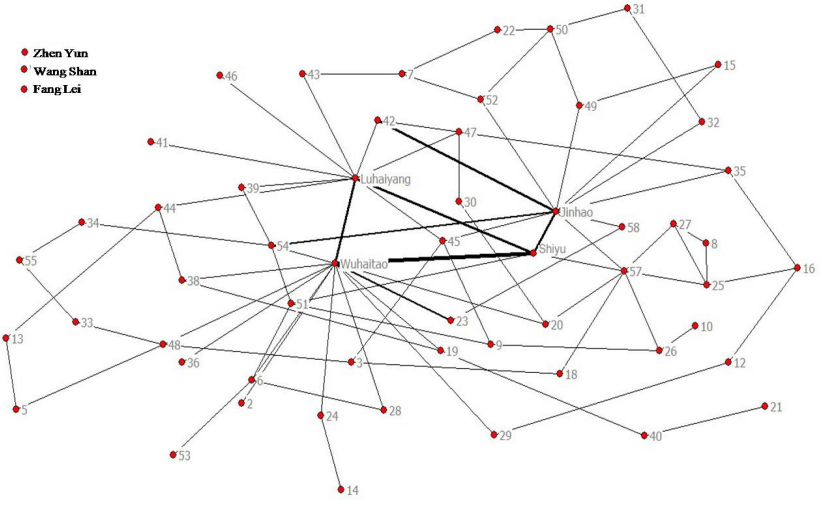


Fig. 1. A Social Network Analytical Graph of 60 Pairs of Parent-Child

From the Figure 1, it can be seen that three or four families are in the key position of the graph, and have formed a team, cooperated, or competed with many other families; While most families are a little bit common, for they have just formed a team with other families for one or two times; Still three families are in a marginal position and have never cooperated with other families. The graph above shows that Lu Haiyang, Wu Haitao, Jin Hao and Shi Yu families belong to the key families, while Zhen Yun, Fang Lei, and Wang Shan families are the marginal ones.

The methodology adopted integrates the analysis of player's parent-child cooperation, the parent-child social network, and the mechanism of Log which records every detailed behavior taking place in cooperation; the cases analyzed cover four families, i.e., Lu Haiyang, Tong Han, Zhen Yun and Fang Lei.

3.1 Analysis of Time Spanning of the Four Pairs of Parent-Child Players

In order to know which kind of behaviors the four pairs of parent-child players conduct, and how much time they spend in these activities, we collect the dynamic data of the players' behaviors and draw a graph to show how much time these pairs of parent and child spend in the game. The following Figure 2 presents a clear model of the four pairs of parent-child players' behaviors. Axis Y represents the total time spanning, and axis X represents the players' behaviors:

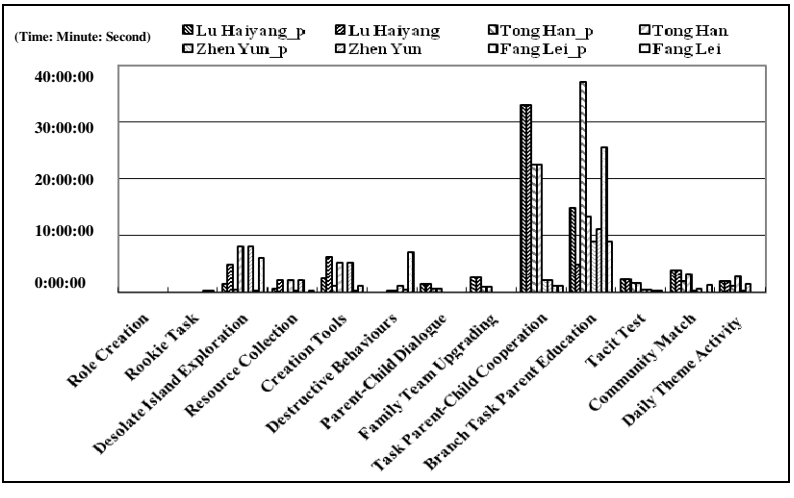


Fig. 2. Time Spanning of the Four Pairs of Parent-Child Players

From Figure 2 it can be easily found that the maximum time input by Family Lu Haiyang takes up 1/2 of the total invested in main thread task of the parent-kid cooperation, Family Lu Haiyang also leads the time input of parent-child dialogue, an active cooperation with other families to accomplish the task, creation of more appliances and participation of various activities in community all attribute to their success. The Family Tong Han tops the parent-kid online time input, but that for main thread task of parent-child cooperation accounts for less than 1/3 of the total. Without the help from parents, even the adolescent player puts in more time, it is comparatively difficult for him to control two players and accomplish the task in the meantime. In this way, more time only yields to a low success ratio, which leads to a near null growth of intimacy; therefore, player's entrance to a higher level is deterred. Parents put most of their time in achieving branch thread task in the experiencing section of parental education. The Family Zhen Yun's total time online is not long, but they spend 1/2 of their time online in cooperation, which shows an active

communication between parents and kids. The adolescent spend much time in creating tools and accomplishing branch thread task in the experiencing section. The Family Fang Lei has few parent-child communication, the parents' time online as well as the time spent in parent-child cooperation are among the lowest of all families, for the adolescent spend most of time in act of destruction and upgrade of experience so that his family bears no social connection with the other families.

3.2 Analysis of Branch Task Success Ratio of Parent Role in Parental Education

By collecting dynamic data of various parent players' behavior in accomplishing branch thread task of parental education on the server and using cumulative variable mechanism to accumulate numbers of task received and succeeded, we get parents' task process Table 2 in the experiencing sector of parental education:

Table 2. Parental Player's Branch Task Process Report in the Experiencing Sector of Parent Education

Object	Parent Player's Branch Task	Task1: Know One's Own Expectation Task2: Know Children's Aim of Act Task3: Parents' Self-Change Task4: Be Spontaneous to "Listen" and "Think" in Child's Shoes Task5 : Have Right to Question and Express Ideas Task6: Principles and Challenges of Democratic Education Task7: Natural and Logical Result of Act Task8: Wisely Encouraged Task9: Democratic Manner of Upbringing Task10: Family Meeting									
		Task1	Task2	Task3	Task4	Task5	Task6	Task7	Task8	Task9	**Task10**
60 Pairs of Parent-Child Examples	Success Times	**60**	**60**	**60**	**59**	**58**	**59**	**59**	**59**	**58**	**59**
	Attempt Times	**133**	**152**	**149**	**179**	**173**	**153**	**133**	**101**	**162**	**97**
	Success Ratio	**45.11 %**	**39.47 %**	**40.27 %**	**32.96 %**	**33.53 %**	**38.56 %**	**44.36 %**	**58.42 %**	**35.8 %**	**60.82 %**
	Give-up Ratio	**0 %**	**0 %**	**0 %**	**0 %**	**1.67 %**	**1.67 %**	**1.67 %**	**1.67 %**	**3.33 %**	**1.67 %**
Lu Hai yang Family	Success Ratio	**100 %**	**50 %**	**100 %**	**50 %**	**50 %**	**50 %**	**100 %**	**100 %**	**50 %**	**100 %**
Tong Han Family	Success Ratio	**33.33 %**	**25 %**	**50 %**	**25 %**	**16.67 %**	**25 %**	**33.3 %**	**50 %**	**33.3 %**	**50 %**
Zhen Yun Family	Success Ratio	**50 %**	**33.3 %**	**50 %**	**25 %**	—	—	—	—	—	—
Fang Lei Family	Success Ratio	**16.67 %**	**14.29 %**	**12.5 %**	**0 %**	**0 %**	**12.5 %**	**14.29 %**	**25 %**	—	**12.5 %**

From Table 2, it can be learnt that even substantial rewards do not lead to branch task necessarily, but most parents will do their utmost to finish it. 0% Give-up ratio for branch thread task is a strong proof of parents' interests and impetus in parental education. However, their low success ratio proves that parents who live under a traditional family background learn and inherit last-generation's *teaching method* instead of the new one. Therefore, whilst dealing with the new generation living in a modern and democratic world, they lose themselves in finding an effective way to educate their children. Special attentions should be given to *spontaneous to listen and think in child's shoes, Having Rights to question and express ideas and Democratic manner of upbringing*, in which low success ratio explains parents' lack of effective communicative skills while dealing with children.

Performances of the 4 families in accomplishing branch thread task of parent education turn out to be quite different: parents of Family Lu Haiyang has a relatively high success ratio in accomplishing tasks, which proves their efficiency in dealing with parent-child conflicts and their family education knowledge in communicating with children and controlling their emotions. Parents of Tong Han family may fluctuate around the average level when accomplishing their tasks, but they spend much time to actively learn communicative skills and family education knowledge as a way to becoming efficient parents and bringing up responsible children. Though with a success ratio is beyond average in the early period, parents of Family Zhen Yun choose to give it up later. In an access to user's information in players' account database, it is noticed that parents have not logged onto the game for a long time (During the telephone interview, we find parents are away on their business trip, therefore the game is rarely played owning to busy work). Parents of Family Fang Lei have an extremely low success ratio in accomplishing branch thread task of parent education, some are even 0%, and one branch thread task is even given up.

Our subsequent interviews with parents of four case families turn out that a lack of guidance is the main reason parents feel confused and anxious during their role conversion period, children's *psychological weaning* period and period when values both old and new clash. The experiencing sector of parental education in this game provides parents with visual perceptual and responsive mode, by means of both positive and passive, this mode teaches parents efficient parent-child communicative skills and emotional controlling methods so that parents can reach *Unity of Theory and Practice* state in practice.

3.3 Traits of Four-Family Samples

The Family Lu Haiyang: it belongs to nuclear family in the online parent-child games; it has cooperated with other families several times. Besides, it participates actively in the community activities, offering help and advice to others. A harmonious parent-child relation, communication and cooperation is beneficial to the healthy development in mentality and physical fitness of teenagers, the formation of their independent personality, the improvement of their studies and their development of peer relations, so that teenagers can grow up healthily in the environment where relationship is intimate and communication is effective.

The Family Tong Han: it belongs to a common family in online parent-child game, with a few experiences in cooperating with other families, parents and kids in this family trust and respect each other, but frictions occur occasionally. Parents' communication with and response to their children is not as frequent as required, which cause unsmooth parent-child cooperation in games. Adolescents are prone to alienate from others, but confused parents often do not know how to guide their child to form active and healthy personality such as self-respect, self-confidence, responsibility etc. In educating adolescents' attitude towards setbacks and difficulties, parents are also short of practical experience.

The Family Zhen Run: it belongs to a special family in online parent-child game, in this family, parents are busy with work and have no time for their child, and therefore parent-child relation will be in danger with the decreasing of communication.

The Family Fang Lei: it belongs to a marginal family in the online parent-child game with little experience in cooperating with other families, lacking of understanding and alienation between parents and kids with a dangerous parent-child relationship. Besides, behaviors like communicating with kids only during mealtime and using commanding tones to teach, together with parents' failure in items like *spontaneous to listen and think in child's shoes, having rights to question and express ideas and democratic manner of upbringing* are all signals that parents lack in parental education knowledge and effective communicative skills. Children growing up in this kind of family will display antisocial and destructive behaviors like drinking, committing crimes, etc. Consequently, teenagers are becoming increasingly sensitive to parents' one-way authority; therefore, they strive for more equality and freedom.

4 Investigation and Conclusion

An FACESII-CV Test is arranged specifically for the 4 given families above. This chart, worked out by Olson, etc. (1982) with its Chinese version revised by Li Pengfei et al. (1991) [12], has high reliability and efficiency. It is a self-evaluating chart with two sub-charts and a total of 30 items. It focuses mainly on two aspects of the family functions: Intimacy that is the emotional relation among family members; Adaptability which is the ability for the family to cope with problems according to different family environments and stages of its development. The chart requires the participants to choose from five ranks varying from "No" to "Always", and brings forward two questions: What are their true feelings about the current situation of the family? And what is the idealistic situation for the family in their mind? There are 16 types of families according to different levels of intimacy and adaptability which include 4 balanced types, 4 extreme types and 8 intermediate types [13]. The results of the test on the 4 families are as follows: The Family Lu Haiyang is a balanced type, The Families of Tong Han and Zhen Yun are an intermediate type and Fang Lei's

family is an extreme type. If conditions allowed, intermediate families can be transformed into balanced families by balancing relations of family members and their adaptation to each other as a way to optimize their family functions. The parent-kid family type reflected through their performance in the game is in line with the test results.

Families can be divided into following types according to their main traits of their behaviors in the parent-kid game:

4.1 "Harmonious" Parent-Child Relationship Families

The Family Lu Haiyang is one of the representatives of the balanced family type. Regarding intimacy, family members in this kind of family keep a good emotional connection while not relying on each other overly. They discuss to make decisions and take part in activities together but still have their own social circles and activities. From the perspective of adaptability, family members share family roles and have a steady division of work and responsibility. Family decisions are always made democratically with compulsive family rules rarely made. Therefore, the family can cooperate in overcoming different kinds of problems and difficulties.

4.2 "Constructive" Parent-Child Relationship Families

The Family Tong Han is one of the representatives of the intermediate family type. If conditions allowed, constructive families can be transformed into harmonious families by balancing relations of family members and their adaptation to each other as a way to optimize their family functions. Parents should learn to understand and adapt themselves to the new system and culture that keep emerging in modern society, while adolescents should fulfill their filial responsibilities and obligations and pursuit the harmonious development by looking after each other.

4.3 "Crisis" Parent-Child Relationship Families

This is an extreme family type with Fang Lei's family being the representative. With reference to intimacy, the family members are too estranged from and indifferent to each other. Each of them does things on his/her own and seldom ask other members for support and help. They also have very few friends or interests beyond the family. In regards to adaptability, their way of handling problems is rigid, embodied in the following two situations: for one thing, one family member has the absolute right to make decisions without consultation; for another, the family has no clear rules and disordered distribution of work, their impulsive decision impulsive is difficult to be carried out. To sum up, "In crisis" family has low rate of intimacy between parents and kids; besides, there are obvious parent-child contradictions, if not solved, can post a threat to adolescents' growth.

5 Inspirations for Designing of the Parent-Child Game

Based upon the discussions above about players' traits in the online parent-child game, we know that the classification of parent-child relationship is still not perfect enough. However, it is definitely important for the designing of parent-child games.

A family that has harmonious parent-child relation usually performs more frequent communications and activities. It takes a very kernel place in the online community. A rich and generous rewarding mechanism is used to encourage the nuclear family to organize online activities more actively, such as, the tacit agreement test between parents and kids, with other families such as the common family and the marginal family. In the virtual game environment, the nuclear family functions as a model to influence other players' online behaviors and helps other families to increase the parent-child intimacy, improves the relationship between parents and kids and enhances the parent-child communication skills. Meanwhile, these nuclear families receive considerable economic benefit, rare resources and various special equipments. The nuclear family can also be arranged to act as the supervisor of the online community and an opinion leader, providing supports to other families. Therefore, the nuclear family can actively organize parent-child activities both online and in real life to create a platform connecting each other both online and offline. A blockade system [14] can be set up in the parent-child game. With the improvement of parent-kid cooperation players can receive many rewards. If accumulated to a certain quantity, these rewards can be used to unlock more special cooperative models. Therefore, it can improve the players' coherent online experience and optimize the relationship between parents and kids. As a result, parents' knowledge in how to educate a child is enhanced and a deeper parent-kid relationship and communication can be achieved.

By contrast, the constructive family has only one or two experiences in cooperating with other families. As a common family in the online community, the developing family, if helped by a nuclear family, can receive rewards much more times, which is a great push for them to become more active in taking part in the community activities and other activities. By their participation, the constructive family can improve the relationship among children; promote the sharing of parental knowledge and experience. By borrowing better communication skills from the nuclear family, the constructive family can enhance their own ones and ensure the relationship between parents and kids in an obviously equal way. Meanwhile, the system can be interfered by setting up progress mechanism that divides a main task into a number of branch tasks in order to continuously accumulate achievements and accomplish tasks. In this way, the family cohesion between parents and kids is enhanced. The system can dynamically adjust the difficulty of tasks as a way to establish a gradual game process that allows parent-child players to grow up. It is necessary to point out that parents should learn to understand and adapt themselves to the new system and culture that keep emerging in modern society while adolescents should fulfill their filial responsibilities and obligations and pursuit the harmonious development by looking after each other. If conditions allowed, constructive families can be transformed into harmonious families by balancing relations of family members and their adaptation to each other as a way to realize a harmonious parent-child relation.

The "crisis" family has few contacts with other families. Gradually, this kind of family is edged aside, its intimacy rate between parents and kids slides, and then problems such as poor communication and inharmonious cooperation emerge. In response to this situation, the system enhances control and supervision over the family, for instance, to establish a rich and generous awarding mechanism, a rational community system and a recurrent and diverse task mechanism, so as to largely ensure the cohesion of players. A ranking list is designed for each task in the community to encourage this kind of family to try to accomplish the same task with a best grade. By doing this, the cooperation and intimacy among parents and kids can be highly improved. More branch thread tasks can be designed in the experiencing sector of parent education. The dialogue system using branch-style mutual color matching mechanism can be used to imitate parent-child relationship between the NPC and parents and program of becoming efficient parents is planted so that the parent's skills and abilities in teaching kids can be enhanced and their knowledge of children's growth, adaptation and development can be widened. Consequently, a much more intimate parent-kid relationship can be built up.

In the further research, more schools will be included and a larger group of parent-child game players will be formed. By collecting and analyzing more dynamic statistics and by finding rules and analyzing results from player's complicated psychology and behaviors, we will strive to make our research more representative.

References

1. China Population Publicity and Education Center: The Report of Teenagers' Healthy Personality surveyed in 2010 (March 13, 2011) http://edu.sina.com.cn/zxx/2011-03-13/1410287962.shtml
2. Hollingworth, L.S.: The psychology of the adolescent. D. Appleton and Company, New York (1928)
3. Collins, W.A., Laursen, B.: Changing relationships, changing youth: Interpersonal contexts of adolescent development. The Journal of Early Adolescence 24, 55–62 (2004)
4. Duncan, L.G., Coatsworth, J.D., Greenberg, M.T.: A Model of Mindful Parenting: Implications for Parent–Child Relationships and Prevention Research. Clinical Child and Family Psychology Review 12, 255–270 (2009)
5. Asai, K., Kondo, T., Mizuki, A., Billinghurst, M.: Lunar Surface Collaborative Browsing System for Science Museum Exhibitions. In: Pan, Z., Cheok, A.D., Müller, W., Zhang, X., Wong, K. (eds.) Transactions on Edutainment IV. LNCS, vol. 6250, pp. 34–43. Springer, Heidelberg (2010)
6. Fidas, C., Komis, V., Avouris, N.: Heterogeneity of learning material in synchronous computer-supported collaborative modelling. Computers & Education 44(2), 135–154 (2005)
7. Wendel, V., Babarinow, M., Hörl, T., Kolmogorov, S., Göbel, S., Steinmetz, R.: Woodment: Web-Based Collaborative Multiplayer Serious Game. In: Pan, Z., Cheok, A.D., Müller, W., Zhang, X., Wong, K. (eds.) Transactions on Edutainment IV. LNCS, vol. 6250, pp. 68–78. Springer, Heidelberg (2010)
8. Yin, R.K.: Case study research: design and methods, 3rd edn. Sage Publications, Thousand Oaks (2003)

9. Merriam, S.B.: Qualitaivative research and case study applications in education, 2nd edn. Jossey- Bass Publishers, San Francisco (1998)
10. Stake, R.E.: The art of case study research. Sage Publications, Thousand Oaks (1995)
11. Chen, X.: Qualitative Research Methods and Social Science Research. Education and Science Press, Beijing (2000) (in Chinese)
12. Wang, X.: Rating Scales for Mental Health. Chinese Mental Health Publications, Beijing (1999) (in Chinese)
13. Olson, D.H., McCubbin, H.I., Barnes, H., Larsen, A., Muxen, M., Wilson, M.: Families: what makes them work, 2nd edn. Sage, Los Angeles (1989)
14. Dunniway, T., Novak, J.: Game Development Essentials: Gameplay Mechanics. Delmar Cengage Learning, New York (2008)

Towards an Open Source Game Engine for Teaching and Research

Florian Berger and Wolfgang Müller

University of Education Weingarten, Kirchplatz 2
88250 Weingarten, Germany
{berger,mueller}@md-phw.de

Abstract. When introducing students without in-depth programming experience to the development of educational games, an off-the-shelf game engine will not fit, especially when users are meant to directly interact with the underlying framework. To accommodate this scenario and still be able to provide state-of-the-art features, we are developing the custom game engine "Fabula". A software stack of Python, Pygame and Twisted ensures an accessible, yet powerful application core, undemanding cross-platform visuals and field-tested networking. Fabula's game world abstraction aims to be intuitive to people who have not been involved in game development before, while at the same time being general enough to fit several genres. The engine's goal to help students explore the creation of enjoyment, emotional responses and social experiences in a game context, keeping the hassle with technical details at a low level, will be evaluated using the Technology Acceptance Model (TAM).

Keywords: educational games, game engine, edutainment, cross-platform, game development.

1 Introduction

While the intricacies and mechanisms of game-based learning are a topic of ongoing research, there by now is a consensus that computer games can be a valuable educational tool. [5] [10] However, while discussing how *playing* educational games can motivate learners or foster cognitive flexibility, it is often overlooked that the *creation* of digital games can be highly instructive.

Kafai et al. have described how designing games for learning mathematics can unlock rarely used informal knowledge both in teachers and learners as early as 1998. [7] Games creation as a means of teaching has been reported as worthwhile in computer science [11] and software engineering education [2]. Renowned universities like the ETH Zurich have adopted game development in their regular undergraduate computer science courses for years. [1]

Understanding how games work and how to create them has recently become important for disciplines other than computer science as well. At the University of

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 69–76, 2012.
© Springer-Verlag Berlin Heidelberg 2012

Education Weingarten, Germany, our team introduces students of Media and Education Management to the development of edutainment applications, especially educational games, following a “Learning by Design” approach.

In doing so, an important technical decision was which game engine to use. A game engine manages the game state and renders the game's audiovisual representation, therefore defining the basic capabilities of the game, the platforms it will run on, and hardware requirements.

In the case at hand an off-the-shelf engine would not fit. If end users without in-depth programming experience are meant to directly interact with the underlying framework, accessibility and instructive design come into play. This paper presents a custom-made software that comprehensively addresses these issues.

2 Constraints

We were faced with the task to select, adapt or create a game engine that would enable students without an extensive IT background to edit, understand and develop games. Following a top-down approach, they should first play and analyze some given examples and subsequently customize media and game logic using a graphical user interface (GUI). Therefore the first demand was to expose low-level game elements in a way that could be used by people without previous programming skills.

Advanced students should optionally be able to work at the code level, which called for a *low-threshold implementation*. Yet the engine had to provide some state-of-the-art features: multi-player games, networking, 2D and possibly 3D visuals. Besides it should be open for later additions, demanding a *well-engineered base*.

Writing software to be used in a research and education environment means dealing with multiple platforms, possibly dated hardware and operating systems and budget constraints in general. The final demand was a *versatile implementation* with a *low cost of ownership* and simple licensing terms.

3 The Software Stack

The latter request naturally called for a look at open source software. Together with the need for straightforward access to the engine details, this ruled out a lot of commercial products: they predominantly rely on a C++-like scripting interface that is unsuitable for novice programmers.

There are a few open source projects that aim to teach multimedia programming to beginners. In *Scratch* [14], developed at the MIT Media Laboratory, users combine given actions in a drag and drop-fashion and link the resulting scripts to 2D graphic objects. *Alice* [12], a software from the Carnegie Mellon University, has a similar concept, but offers more fine-grained scripting and a 3D environment. Both tools are based on the *Squeak* system, which is a combination of the Smalltalk programming language and an object-oriented audiovisual environment.

Scratch and *Alice* aim at teaching programming constructs and syntax instead of game design, and games are just one possible outcome. They also lack a game world abstraction, a server component for multiplayer games and adventure game building blocks like maps and inventories and hence were not suitable for teaching purposes in our given context.

A large part of the remaining open source game engines are in fact 3D rendering engines that also lack an abstraction of a game world. The only product that was left for consideration at that point was the *Arianne Framework* [8], created by Miguel Angel Blanch Lardin and written in Java. It consists of a multi-threaded server, a network protocol on top of TCP, a persistent database and a reference client that implements a 2D role-playing game. Its simple architecture looked promising, but since the Java server did not meet the requirement for accessibility, we decided to use Arianne as a design inspiration, but to write a complete new game engine "Fabula" in line with the demands above.

Programming languages that are known to be easy to learn and handle for novice users include *Lua*, *Ruby*, and *Python*. Among these, Python stands out because of its comprehensive standard library which would allow for an easy addition of new features. Python also provides cross-platform compatibility and therefore became the language of choice for Fabula.

A frequent objection to Python for game development is the runtime execution speed. Python programs are compiled into bytecode and then executed by a virtual machine, an architecture that clearly does not suggest itself for complex real-time rendering. But first, Python can call precompiled C/C++ code, making it usable as the host of a 3D engine as done in the *Panda 3D* software by Disney and the Carnegie Mellon University [6]. And second, the demand to support dated hardware had put restrictions on the audiovisual complexity: for the time being, the games to be developed were planned to feature 2D graphics without hardware acceleration and only sporadic user interaction like in classic adventure or strategy games.

Having selected Python as the development platform and 2D game visuals, the choice of the actual graphic and display module was *Pygame*, a very popular game creation framework for Python [9]. It is a rather thin wrapper around *Simple DirectMedia Layer* (*SDL*), a cross-platform multimedia library. [13] Pygame provides classes and functions to setup a screen, to display and control bitmap objects and to capture user input.

The Python standard library offers networking resources of various abstraction levels. Building upon them, the *Twisted* framework implements more sophisticated asynchronous network servers and clients [4], with features like automatic reconnection and SSL encryption. Since having these already in place would allow more time to work at the actual core of the engine, we chose Twisted to provide all client and server side networking.

The software stack of Python, Pygame and Twisted ensures an accessible, yet powerful application core, undemanding cross-platform 2D visuals and field-tested networking. Still, students should be able to create games without being exposed to that technical layer, conveniently accessing the built- in game world instead.

4 Game World Abstraction

To make Fabula usable for teaching purposes, the abstraction should cover not only a specific game design, but allow for a variety of games to be made, including games with 3D visuals in the future. The main abstractions are *Rooms* and *Entities*.

A *Room* represents an enclosed location. Its data structure is a graph of discrete positions, currently implemented as a two-dimensional grid of squares. Each position in the Room is represented by a *Tile* object, whose type can be `FLOOR` for walkable positions or `OBSTACLE` for non-walkable positions. A game can have an arbitrary number of rooms.

An *Entity* object represents a virtual character or an item. One single character per client is controlled by the player, other characters are non-player characters (*NPCs*) and are either controlled by the server or by players on remote machines. The remaining Entities are *items* that characters can interact with. Each Entity resides on a node in the Room's graph. (Figure 1)

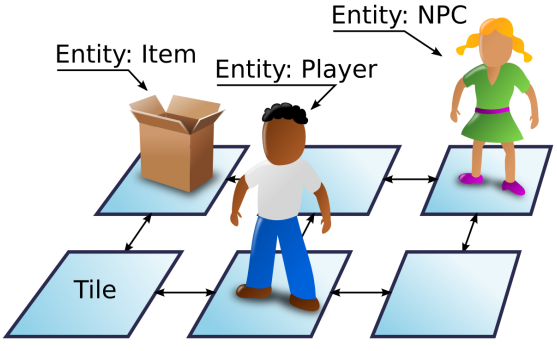


Fig. 1. Fabula game world abstraction: room, tiles, entities

The engine defines a series of possible actions that a player can take. To accommodate players' expectations, these are derived from classical adventure games: player characters can

- move around the room
- look at and manipulate items
- pick up items, in which case they are transferred to an *inventory*
- drop items from the inventory on tiles, other items and NPCs
- talk to other characters

This form of player interaction has been made popular by LucasArts in the late 1980s, and is still being used in contemporary commercial games.

The game world abstraction aims to be intuitive to people who have not been involved in game development before, while at the same time it is general enough to fit several genres—adventures, strategy games, role-playing games, or even business

simulation games. It also allows 2D and 3D games to be built upon it. Together with the networking capabilities this also allows for interesting cooperative games being made.

5 Architecture and Component Design

Fabula comprises a client and a server part. They communicate by sending *Events*. Multiple Events are bundled into *Messages*. Events are implemented as Python objects and contain the relevant data as attributes. Each possible player action is associated with a predefined Event: *TriesToMoveEvent*, *TriesToLookAtEvent* etc., which inherit from an *AttemptEvent* class. The server reacts by sending an instance of *ConfirmEvent* if appropriate, for example *PicksUpEvent* or *MovesToEvent*. These indicate that the server has already changed the game state and that the client has to update its copy and present the Event to the player. There are also Events to spawn or delete Entities, change Tiles or enter a new room.

To simplify the design we are aiming at a symmetrical layout of client and server. Each part consists of an *Engine*, a *Plugin* and an *Interface*.

The *Interface* encapsulates the network transmission of messages using a simple "grab" and "send" API. There are currently two implementations of *Interface*: one using the *Twisted* framework for TCP networking, and one hardwiring Client and Server Interface on the same machine for local single player games. The server handles multiple clients by allocating a *MessageBuffer* for each client in the Server Interface. The server main loop processes client messages round robin, broadcasting the result to all clients in the same room if applicable.

An *Engine* manages the game state, receives messages from and sends messages to the *Interface* and calls the *Plugin* on a regular base. The game state of the *Server* Engine is authoritative, while the *Client* Engine updates a local copy that is the source of the game's presentation.

A *Plugin* serves different purposes in each part. The *Serverside Plugin* implements the game logic. All player attempts run through basic sanity checks in the Server Engine—for example whether the player character is allowed to move to a certain position or can drop a certain item—and, if approved, arrive in the Plugin. For each game a developer must implement a custom logic. Fabula comes with an intuitive GUI editor that allows the user to build a simple condition-response-logic. By executing player actions, for example trying to manipulate an item, the user triggers a look-up of an already defined matching response message. If none is found, the user will be prompted to create a new one. The resulting game logic is saved to a configuration file.

The client-side Plugin implements the *User Interface*, the actual graphical rendering of the game and the point of interaction with the user. There is no obligation how to present the abstract game state to the user and how to collect the input. During the course of development we have written an ASCII-art command line user interface for testing, while the current one is based on Pygame and uses 2D graphics. Because of the general design, both can be used side by side.

Figure 2 shows a chart of the Fabula architecture.

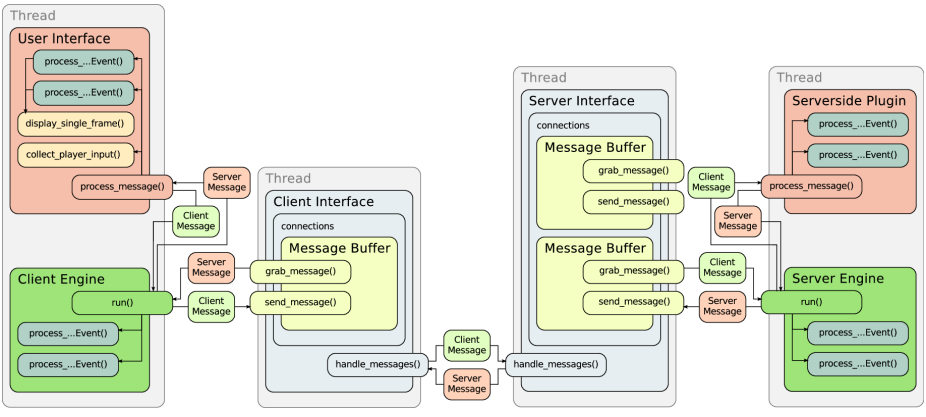


Fig. 2. A chart of the Fabula game engine architecture

The basic game engine objects—*Room*, *Entity*, *Engine*, and *Plugin*—inherit from a common base class and hence expose the same API. This makes it easy for developers to customize their behaviour.

In addition, Fabula provides an *Asset* manager to fetch and cache media files—2D graphics, sound, 3D models—from the local file system or over the network using an URI. Since it is presentation agnostic, the code makes no assumptions about the data being fetched and simply makes them available to the User Interface.

6 Future Work: Extension and Evaluation

The first public version of Fabula has been released in January 2011 under the General Public License (GPL). Server, Client, their Plugins and the Interfaces are implemented and working as shown in figure 2. A Pygame-based user interface is in place, and there is a first version of the graphical editor that incorporates the aforementioned game logic editing (figure 3). Both Client and Server run on current and older versions of GNU/Linux and Microsoft Windows.

As a test case, a serious game teaching project management skills is being developed alongside the engine. The upcoming versions will see more features being added to Fabula, such as an OS X port, scaling for multiple CPUs and support for mobile devices.

As laid out in the intruduction, our team at the University of Education Weingarten will use Fabula for student projects in the “Media and Education Management” curriculum. These students, not having an extensive computer science background, shall first get accustomed to handling game assets and logic, then get to know the basic game design iteration of prototyping and play-testing. Finally they shall be able to create a complete edutainment application over the course of their project. Given

Fig. 3. Screenshot of the work-in-progress version of the Fabula game editor

their lack of programming skills, their acceptance of the Fabula system is a critical factor for the success of this educational setting.

Instructor characteristics as "teaching style", *student* characteristics as "computer competency" or "interactive collaboration", *technology* and *support* have been reported as success factors of e-learning acceptance [15] and appear to be applicable in our case as well. The Technology Acceptance Model (TAM) by Davis et al. provides even more detailed means for measuring and monitoring the *perceived usefulness* and the *perceived ease of use* of a computer system. [3] TAM has been extended to incorporate *control*, *intrinsic motivation* and *emotion* by Venkatesh [16], an addition that we view as highly suitable for a game-related context. We plan to use this established theoretical framework alongside the introduction of Fabula to evaluate how the software helps students to explore the creation of enjoyment, emotional responses and social experiences in a game context, while trying to keep the hassle with technical details at a low level.

Acknowledgments. A year of funding for Fabula has been granted by the MFG Foundation of Baden-Württemberg, Germany, in October 2010.

References

1. Bay, T., Pedroni, M., Meyer, B.: By students, for students: a production-quality multimedia library and its application to game-based teaching. Journal of Object Technology 7(1), 147–159 (2008), http://www.jot.fm/issues/issue_2008_01/article5/

2. Claypool, K., Claypool, M.: Teaching software engineering through game design. In: Proceedings of the 10th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education (ITiCSE 2005), pp. 123–127. ACM, New York (2005)
3. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 35(8), 982–1003 (1989)
4. Fettig, A.: Twisted Network Programming Essentials. O'Reilly Media (2005)
5. Foreman, J.: Game-Based Learning: How to Delight and Instruct in the 21st Century. EDUCAUSE Review 39(5), 50–66 (2004)
6. Goslin, M., Mine, M.R.: The Panda3D graphics engine. IEEE Computer 37(10), 112–114 (2004)
7. Kafai, Y.B., et al.: Game Design as an Interactive Learning Environment for Fostering Students' and Teachers' Mathematical Inquiry. International Journal of Computers for Mathematical Learning 3(2), 149–184 (1998)
8. Lardin, M.A.B., Russell, K.: Arianne—an open source multiplayer online framework to easily create games (2011), `http://arianne.sourceforge.net/about.html` (retrieved on April 20, 2011)
9. McGugan, W.: Beginning Game Development with Python and Pygame: From Novice to Professional. Apress (2007)
10. Mishra, P., Foster, A.: The Claims of Games: A Comprehensive Review and Directions for Future Research. In: Carlsen, R., et al. (eds.) Proceedings of Society for Information Technology & Teacher Education International Conference 2007, pp. 2227–2232 (2007)
11. Overmars, M.: Teaching computer science through game design. IEEE Computer 37(4), 81–83 (2004)
12. Pausch, R., et al.: Alice: Rapid Prototyping System for Virtual Reality. IEEE Computer Graphics and Applications (May 1995)
13. Pendleton, B.: Game programming with the sdl. Linux Journal 2003(110) (2003)
14. Resnick, M., et al.: Scratch: programming for all. Communications of the ACM 52(11) (2009)
15. Selim, H.M.: Critical success factors for e-learning acceptance: Confirmatory factor models. Computers & Education 49(2), 396–413 (2007)
16. Venkatesh, V.: Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. Information Systems Research 11(4), 342–365 (2000)

Designing a Mixed Digital Signage and Multi-touch Interaction for Social Learning

Long-Chyr Chang and Heien-Kun Chiang

Information Management Department and Graduate Institue of Digital Content Technology & Management, National Changhua University of Education, Taiwan
{lcchang,hkchiang}@cc.ncue.edu.tw

Abstract. The use of digital signage for informal learning on campus remains a less unexplored research area. In this paper, we present a mixed digital signage and multi-touch interaction environment that encourages students to engage in a watch-and-play style for social learning at a public place. To realize the key factors influencing students' engagement and satisfaction with our system, we developed a user study model based on the updated D&M IS Success Model and the Technology Acceptance Model. The integrated model contains six constructs: system quality, information quality, interaction quality, perceived usefulness, perceived ease-of-use, and user satisfaction. Empirical result indicates the proposed system is very promising. However, user study on our research model also validated some system design problems that must be solved to make the proposed system a great learning aid in the foreseeable future. The findings of this study can provide useful guidelines for researchers interested in developing advanced social learning systems.

Keywords: Multi-touch, Digital Signage, Social Learning.

1 Introduction

Recently, digital signage has been getting popular in public and private environments, such as retail stores, museums, corporate buildings and even campuses. Digital signage with large and vivid LCD displays is typically used as a public information exhibition or "narrowcasting" to show information, advertisement, and other messages. Today, modern zoos, parks and museums usually use such facility to serve as a visitor information system to attract their visitors [8]. Using digital signage as an informal learning tool is not a new idea. In spite of the growing popularity of digital signage, the use of digital signage as an informal learning tool on campus remains a less unexplored research area because it involves the complexity of developing new hardware and software innovations. The lack of interactive capabilities of conventional digital signage prevents students from effectively exploring and sharing information on large displays for learning purpose. Even though interactive digital signage is commercially available now, it has yet to become a common commodity on most campuses.

Multi-touch interaction with large display, on the other hand, provides natural and playful interfaces that can potentially attract students to engage and play in public

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 77–87, 2012.
© Springer-Verlag Berlin Heidelberg 2012

space. Ever since Han published and demonstrated his sensational multi-touch presentation at the 2006 TED (Technology Entertainment Design) conference [9-10], the multi-touch technology has been gaining great attention worldwide. The introduction of Apple's iPhone in January 2007 and recently iPad and Windows 7 has spurred the enthusiasm from general public, and propelled more researches into the multi-touch sensing technology.

In this paper, we present a mixed digital signage with multi-touch interaction interface that encourages students to engage in a watch-and-play style for social learning at a public place. To make the system more fun and playful, it was designed as a Web-based system with the rich media characteristics of Flash and the natural interaction capabilities of multi-touch technology. The goal of this study is to investigate how to design a natural and playful interface for conventional large displays so that it not only can be used for digital signage but it also can attract students to engage in a social learning space. Students can explore learning materials of computer animation and computer game, interact with their peers in small groups, perform creative brain-storming activities collaboratively, or showcase their animation works or game works outside the classroom. To realize the key factors influencing students' engagement and satisfaction with our system, we developed a user study model based on the updated D&M IS Success Model [6] and the Technology Acceptance Model (TAM) [4]. The integrated model contains six constructs: system quality, information quality, interaction quality, perceived usefulness, perceived ease-of-use, and user satisfaction. The prototype system was set up for weeks outside the department office to let students to experiment and play. Empirical study using the survey questionnaires from participating students indicates the proposed system is promising. The findings of this study can provide useful guidelines and implication for researchers interested in developing advanced social learning systems.

2 Related Works

Multi-touch large display is attractive yet expensive. We are interested in the design and evaluation of a mixed digital signage and multi-touch interface for social learning. To this end, we review previous researches of multi-touch large displays and evaluation models of e-learning systems.

2.1 Multi-touch Large Display Applications

Digital signage and multi-touch interaction have attracted considerable attention in recent years because they provide a novel yet effective way of human-computer communication in many public spaces. For example, NEC's Eye Flavor utilized digital signage with an embedded camera to gather demographics information about visitors for direct targeted advertising in 2008 [13]. Hosia (2010) installed a large public display in Oulu urban city to allow visitors to explore city information [12]. In fact, using inexpensive digital signage to support the cooperative sharing and exchange of information is not new. The Dynamo system used two 50-inch plasma screens positioned side-by-side in a public space to enable students to meet

occasionally for gathering, displaying, sharing, and exchanging a wide variety of information [2]. The ease-of-use and direct manipulation nature of interactive gesture interface in multi-touch device have great potential for supporting multi-user collaboration on large screens. For example, Microsoft's Surface Table can support multi-touch learning experience allowing a group of students to work simultaneously on its surface [15]. The CityWall project [17], building a large multi-touch display wall in a central location in Helsinki, provides a collaborative, playful and social learning interface for visitors to participate the discussion and interaction in the ever-changing city landscape. The potential usage of multi-touch interaction in education is becoming recognized; however, how to effectively use such technology for formal or informal learning is still an active research area.

2.2 Evaluation Model of e-Learning Systems

To realize the key factors influencing students' engagement and satisfaction with the proposed system, we need to develop a model for user study. Although the proposed new Web-based system, mixed with digital signage functions and multi-touch applications, is designed as a new social learning environment, it basically can be viewed as a Web-based e-learning system. Two well-known models for studying user behavior and success in information systems, the Technology Acceptance Model (TAM) [4] and Delone & Mclean's updated IS Success Model [5-6], are employed as the basis of this user study. TAM model considers perceived usefulness and perceived ease-of-use as two primary factors influencing users' acceptance of a new computer technology. Perceived usefulness, defined as "the degree to which a person believes that using a particular system would enhance his or her job performance", is a major determinant of users' intentions to use computers; while perceived ease-of-use, "the degree to which a person believes that using a particular system would be free from effort", is the secondary determinant of users' intentions [4].

The TAM theory provides an important guideline for us in designing a new computer application. It implied that although the multi-touch interaction's ease-of-use can provide a natural and ease-of-use interface, the usefulness of our system is more important in fulfilling users' needs. The widely used success model developed by Delone and Mclean in 1992 provides a framework for identifying and measuring the relative impact and causal associations of various constructs that leads to system success [5]. Ten years later, DeLone and McLean introduced an updated construct, service quality, to their IS success model [6]. Now, the model consists of six interrelated constructs: information, system, service quality, (intention to) use, user satisfaction, and net benefits to gauge if an information system is successful. To evaluate the success of a system, one can measure its information, system, and service quality constructs whose characteristics will affect users' intension to use the system. The net benefits construct indicates positive or negative impact on users' satisfaction and their further use of the information system.

Later, the D&M model was used to evaluate the success of e-commerce [7] and e-learning systems [20, 14, 11]. Wang et al. (2007) validated the D&M model and concluded that it can be used to assess the success of e-learning systems in the context of an organization [20]. Lin et al. (2007) studied how system, information and service quality influence the successful use of a Web-based learning system [14]. In their

study, e-learning was seen as a service by a private organization. Holsapple and Lee-Post (2006) adapted D&M to the e-Learning Success Model, which posits that the overall success of an e-learning initiative depends on the attainment of success at each of the three stages of e-learning systems development: system design, system delivery, and system outcome [11].

3 The Mixed Digital Signage and Multi-touch Interaction System

We designed and developed a mixed digital signage and multi-touch interface, as shown in Figure 1, by transforming two VIZIO 52-inch LCD displays into a multi-touch social learning space using image processing technique and rear DI (diffused illumination) method [19].

(a)The interactive digital signage interface

(b)Performing multi-touch interaction

(c)Multi-touch applications

(d)Students' multimedia works

Fig. 1. The multi-touch digital signage learning space

The CCV (Community Core Vision) [3] software was employed to compute the touched positions and then transfer them to the application programs via TUIO protocol. The proposed system can detect up to 10 touched positions. The two LCD displays are designed to function as a typical digital signage, placed just outside our department office, to attract students' interest and intention.

The watch-and-play style of the system enables students to watch multimedia learning video, and to demo their animation work or game works to other students

with their both hands. Students are encouraged to watch other students' activities such as playing games, solving problems, or demonstrating their design or art skills. From the observation of others' actions, they then can gain the idea on how to interact with the system and try to perform similar actions to solve their own problems. Figure 1a shows the interactive digital signage interface. A hot spot indicates a touched area that transforms the signage interface into multi-touch applications. Figure 1b shows students' operations on multi-touch interaction. Figure 1c shows the interface of multi-touch applications, and figure 1d demonstrates students' multimedia works.

3.1 System Architecture

The proposed learning system provides innovative interfaces with mixed digital signage and multi-touch interaction to promote students' social learning willingness on the computer and animation course. The system architecture is shown in Figure 2.

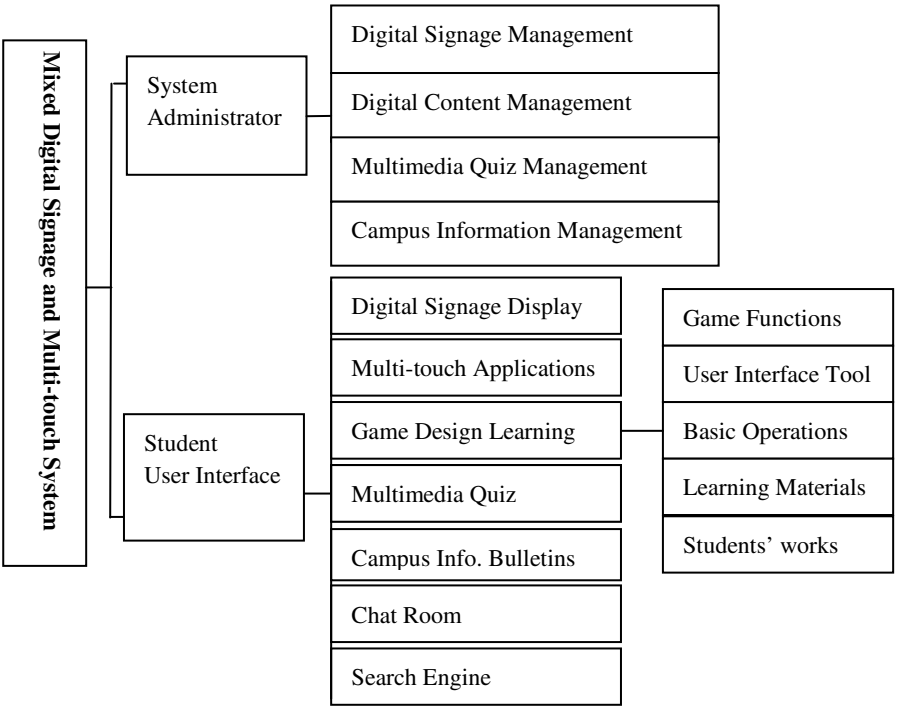


Fig. 2. The system architecture

To provide a playful, multimedia-enriched, and interactive learning space, the system is designed as a Web-based client-server application. The Web server runs on top of Ubuntu Linux operating system and MySQL database system. It also uses XML for data description and PHP as server-side language to return Flash-embedded

HTML content to its clients. The web server side contains the following components: Digital Signage Management System, Digital Content Management System, Multimedia Quiz Management System, and Campus Information Management System. The Digital Signage Management system schedules multimedia information such as texts, animations, video information and broadcast them to the client side according to the course schedule. Digital Content Management System manages students' works on all digital design classes such as 3D animation, Photoshop, and game courses. The Multimedia Test Management System provides Flash-style multimedia test interface to allow users to interact with quiz questions using their fingers on multi-touch display. The client side interface contains the following components: Digital Signage Interface, Multi-touch Applications, Game Design Learning System, Multimedia Quiz, Campus Information Bulletins, Chat Room, and Search Engine. Furthermore, the Game Design Learning System consists of game functions, user interface tools, basic operations, learning materials and students' works, which are designed to support the watch-and-play style of learning and demonstration. On the client side, the client software is developed using Flash and ActionScript 3.0 for multimedia interaction, and open source Touchlib for multi-touch functions.

4 Evaluation Model and Empirical Study

4.1 The Evaluation Model

We are interested in students' responses in using the watch-and-play style of e-learning system. In addition, how to evaluate the proposed system to see its feasibility and usefulness is also important. The user study model we developed are based on the models developed by D&M [5-7] and Davis [4] as discussed previously. The user study model adopted the information quality, system quality, user satisfaction constructs from D&M model, and it also adopted the perceived usefulness and perceived ease-of-use from Davis's TAM model. In the proposed e-learning system, interaction plays an important role. Therefore, we introduced another construct: the interaction quality construct. As a result, the revised user study model contains six constructs: information quality, system quality, interaction quality, perceived usefulness, perceived ease of use, and user satisfaction. In addition, the measures used for each construct are included in the user study model, as shown in Figure 3.

Among them, information quality, system quality, and interaction quality constructs can be viewed as system design measures. Perceived usefulness and perceived ease-of-use constructs can be thought as attractive measures. The user satisfaction construct can be considered as engagement and satisfaction measures.

To study the causal association of each construct, we also proposed nine hypotheses as shown in the path of the user model. For example, the path H1: System Quality→Perceived Usefulness assumes the null hypothesis that the system quality of our system has positive impact on students' perceived usefulness of using the system; the path H2: System Quality→Perceived Ease-of-Use assumes the null hypothesis that the system quality of our system has positive impact on students' perceived ease-of-use of using the system, and so on.

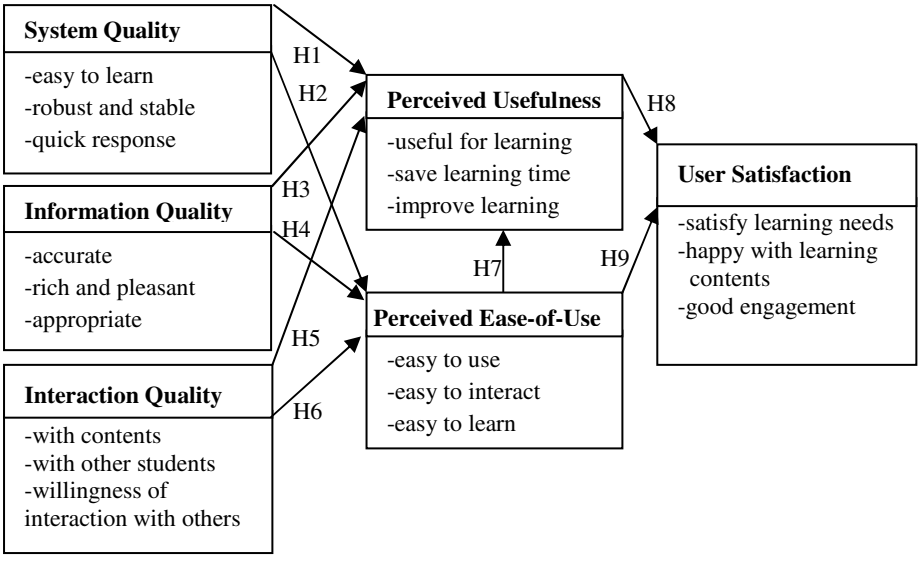


Fig. 3. The user study model for the proposed e-learning system

4.2 The Questionnaire

To realize the participants' multi-touch "engagement experience" and the key factors influencing their satisfactory learning performance using the learning system, the questionnaire was based on the model described in Figure 3. System Quality measures the characteristics such as usability, reliable, easy-to-use, stable, fast, and response time of an e-commerce or e-learning system [7, 11, 20]. We employed "easy-to-learn", "robust and stable" and "quick response" as the system quality measures of the proposed learning system. Information quality captures the learning content issues. We want to know how well the quality of information provided by our multi-touch social learning system. We used "accurate", "rich and pleasant" and "appropriate" as the measures of information quality. Interaction quality was the added construct to measure the usage of multi-touch interaction of the web-based social learning space. Since interaction quality can be thought as a replacement for the service quality in D&M model, we employed interaction quality to measure how well do students interact with the learning contents and how do they interact with each other when using the proposed multi-touch social learning system. We were also interested in knowing whether the system can promote students willingness of interaction with others. Perceived usefulness and perceived ease-of-use are two constructs used by Davis' TAM. We employed the two constructs to measure the attractiveness of the proposed system to students. Therefore, we used "useful learning contents", "save learning time" and "improve learning quality" as the measures for the perceived usefulness, and "easy-to-use" "easy-to-interact" "easy-to-learn" for the perceived ease-of-use. Finally, user satisfaction measures students' satisfaction of using the proposed system. We measures students' satisfaction with "learning needs", "learning contents" and "engagement interaction". The questionnaire contained 18 questions in total and used Likert scale from 1 to 7 for totally dissatisfaction to total satisfaction.

4.3 Empirical Study

Two mixed digital signage and multi-touch interaction systems for social learning are developed and exhibited outside our department office for a few weeks. The watch-and-play style system provides a playful learning space that attracts students to engage and interact with each other. Both questionnaire surveys and video data were collected as part of the field evaluation. The video provides observations of students' interaction with our learning system. Questionnaires are given to participants after they have experimented the proposed system. There are 145 valid questionnaires in total, of which 57% of participants were females, and 43% were males. Regarding "touch" experience, 65% of participants have been experienced with "touch computers, while 35% of participants were not. Statistics software SPSS 17 and SmartPLS 2.0 [18] are used for data analysis. The reliability test of the questionnaire was conducted and the result was shown in Table 1. It showed the questionnaire was consistent because their Cronbach's alpha values were greater than 0.7 on every construct [16].

Table 1. Reliability Test

	System Quality	Information Quality	Interaction Quality	Perceived Usefulness	Perceived Ease-of-Use	User Satisfaction
Cronbach's alpha	0.797	0.808	0.843	0.731	0.726	0.815

The research model was investigated and validated using partial least square (PLS) for structural equation model (SEM) [1]. We use smartPLS 2.0 for path analysis to assess the relationships among constructs in the proposed model. Figure 4 shows the result of using SmartPLS 2.0 with bootstrapping techniques. The path coefficients and *t*-values (two-tailed) are shown for each path of the model in Figure 4. The

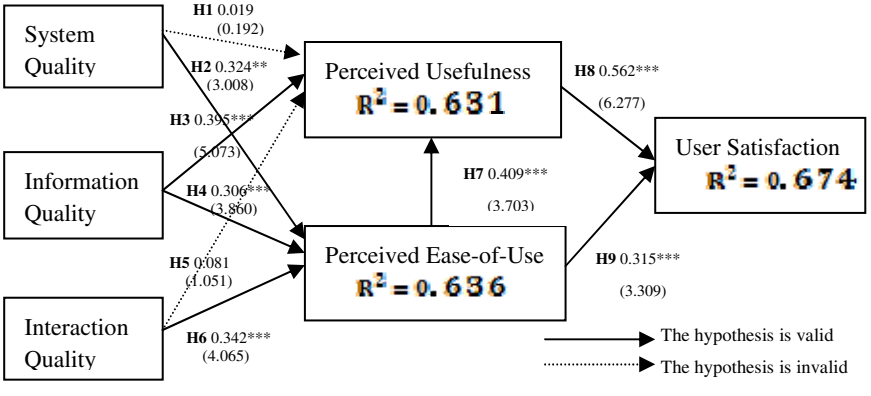


Fig. 4. The structural model and path coefficients

explanatory power of the structural model is evaluated by R^2 value. (Note: * indicates $t > 1.96$ and $p < 0.05$; ** indicates $t > 2.58$, $p < 0.01$; *** indicates $t > 3.29$, $p < 0.001$; the number inside () indicates t value). Table 2 summarizes the hypothesis test of the proposed model.

From the analyses in Figure 4, most null hypotheses are valid except H1 and H5 (see Table 2). The validity of H3 and H4 indicates that information quality of our system has positive impact on perceived usefulness and perceived ease-of-use that make the system attract students to use and learn. The validity of H2 and H6 indicates system and interaction quality has positive impact on perceived ease-of-use. It indicates that our watch-and-play style mixed digital signage and multi-touch system is easy to use, easy to learn, and easy to interact. The validity of H7, H8 and H9 indicates that perceived ease-of-use has positive impact on perceived usefulness and user satisfactory, and perceived usefulness has positive impact on user satisfactory. The explanatory power (R^2) of latent variables: perceived usefulness, perceived ease-of-use, and user satisfactory are 0.631, 0.636, and 0.647 respectively. Factors impacting user satisfaction include system quality, information quality and interaction quality of design construct, and information quality is considered having the most contribution on other constructs.

Table 2. Hypothesis and path coefficients of the model

Hypo.	Path	Path Coeff.	t-value	R^2	Outcomes
H1	System Quality→Perceived Usefulness	0.019	0.192	0.631	invalid
H3	Information Quality→Perceived Usefulness	0.395	5.073		valid
H5	Interaction Quality→Perceived Usefulness	0.081	1.051		invalid
H7	Perceived Ease-of-Use→Perceived Usefulness	0.409	3.703		valid
H2	System Quality→Perceived Ease of Use	0.324	3.008	0.636	valid
H4	Information Quality→Perceived Ease of Use	0.306	3.860		valid
H6	Interaction Quality→Perceived Ease of Use	0.342	4.065		valid
H8	Perceived Usefulness→User Satisfactory	0.562	6.277	0.674	valid
H9	Perceived Ease of Use→ User Satisfactory	0.315	3.309		valid

The invalidity of H1 and H5 indicates the system quality and the interaction quality do not have significant positive effect on perceived usefulness (path coefficient = 0.019, t-value = 0.192; path coefficient = 0.081, t-value = 1.051 respectively). From the findings of interviews with students, they indicated that there are some problems using our system: (1) multi-touch functions are still not quite stable, (2) multi-touch functions lack accuracy in open environment and it can be easily interfered with the infrared rays of the environment, and (3) the response is sometimes not quick enough. This indicates that most students perceived our learning space is useful and ease-to-use in spite of some potential problems in multi-touch stability.

Overall, the user study indicated that our designed system shows very positive results that it can attract students to engage in watch-and-play style social learning and provides satisfactory results. However, there are still some potential problems needed to be solved to make our system more successful. Further interviews with participants, they considered that our system provides innovative interface that is simple and easy to use, learn and interact. They have fun to watch other students

performing multi-touch interaction with learning contents and other students. They thought our systems can promote their willingness to interact with their peers and enhance their learning. If the interaction and system quality can be enhanced then they considered our system would be a great learning aid in the foreseeable future.

5 Conclusion, Implication and Future Research

We designed and developed a mixed digital signage and multi-touch interaction by transforming two VIZIO 52-inch LCD displays into a multi-touch social learning space with little extra cost. The social learning space provides innovative watch-and-play style multi-touch interaction to foster social learning for our computer and animation course. The two LCD displays are designed to function as a typical digital signage to show interesting news and video to attract students outside our department. Students then can point their fingers on displays and turn the system into a multi-touch web-based rich multimedia learning environment. They are able to watch multimedia video, demonstrate their animation or game works to other students with their hands or just watch other students' interaction with the system.

To realize the key factors influencing students' engagement and satisfaction in using our system, we developed a user study model based on the updated D&M IS Success Model and the Technology Acceptance Model. The integrated model contains six constructs: system quality, information quality, interaction quality, perceived usefulness, perceived ease-of-use, and user satisfaction. Empirical result indicated our proposed system is very promising. Most null hypotheses are valid except two hypotheses that indicated the system and interaction quality of the proposed system do not have positive impact on perceived usefulness. Among the above three design factors impacting user satisfaction, information quality is found to contribute most. This indicates that most students perceived our learning space is useful and easy to use in spite of some potential problems in multi-touch stability. Overall, empirical result indicates our solution is very promising. Not only the multi-touch learning space attracts most of our students to use the system, but also students all considered it to be a great learning aid in the foreseeable future. The user study we developed can be used as an evaluation tool to provide useful guidelines for researchers interested in developing advanced social learning systems.

In spite of our effort on developing an innovative mixed digital signage and multi-touch system, our current study only provide a feasible and affordable solution for using such a new technology for social learning. We do not elaborately assess students' learning outcomes because it requires more elaborative and thoughtful study on how to motivate students' learning in such an informal learning space.

As for future research direction, multi-touch social learning space should be made smarter. For example, (1) it should support smart interaction mechanism for users to navigate, explore, communicate and collaborate in 3D virtual world, (2) it should be able to identify students, know where they are, and interact accordingly. Besides, it should be able to provide appropriate learning contents to fulfill students' learning needs.

Acknowledgments. The project is sponsored by National Science Council, Taiwan, with grant# NSC-98-2511-S-018-012 and 99-2511-S-018-018.

References

1. Bagozzi, R.P., Yi, Y.: On the Evaluation of Structural Equation Models. Journal of the Academy of Marketing Science 16(1), 74–94 (1988)
2. Brignull, H., Izadi, S., Fitzpatrick, G.: The Introduction of a Shared Interactive Surface into a Communal Space. In: 2004 ACM Conference on Computer Supported Cooperative Work, pp. 49–58 (2004)
3. CCV, Community Core Vision (2009), http://ccv.nuigroup.com/
4. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Manage. Sci. 35(8), 982–1003 (1989)
5. DeLone, W.H., McLean, E.R.: Information Systems Success: The Quest for the Dependent Variable. Information Systems Research 3(1), 60–95 (1992)
6. DeLone, W.H., McLean, E.R.: The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. J. Manage. Inf. Syst. 19(4), 9–30 (2003)
7. DeLone, W.H., McLean, E.R.: Measuring E-Commerce Success: Applying the DeLone & McLean Information Systems Success Model. International Journal of Electronic Commerce 9(1), 31–47 (2004)
8. Görlitz, G., Schmidt, A.: Digital Signage: Informal Learning in Animal Parks and Zoos. In: Bonk, C., et al. (eds.) Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, pp. 841–846 (2008)
9. Han, J.Y.: Low-cost Multi-touch Sensing through Frustrated Total Internal Reflection. In: Proceedings of UIST, USA, pp. 115–118. ACM Press (2005)
10. Han, J.Y.: TED presentation (2006), http://www.ted.com/speakers/jeff_han.html
11. Holsapple, C.W., Lee-Post, A.: Defining, Assessing, and Promoting E-Learning Success: An Information Systems Perspective. Decision Sciences Journal of Innovative Education 4(1), 67–85 (2006)
12. Hosio, S., Jurmu, M., Kukka, H., Riekki, J., Ojala, T.: Supporting Distributed Private and Public User Interfaces in Urban Environments. In: 11th Workshop on Mobile Computing Systems Applications, pp. 25–30 (2010)
13. Keferl, M., Digital Signage: NEC Eye Flavor Reads Consumers (2008), http://www.japantrends.com/digital-signage-nec-eye-flavor-reads-consumers/
14. Lin, H.-F.: Measuring Online Learning Systems Success: Applying the Updated DeLone and McLean model. CyberPsychology & Behavior 10(6), 817–820 (2007)
15. Microsoft Surface (2007), http://www.microsoft.com/surface/whatissurface.aspx
16. Nunnally, J.C.: Psychometric theory, 2nd edn. McGraw Hill, New York (1978)
17. Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., et al.: It's Mine, Don't Touch!: Interactions at a Large Multi-touch Display in a City Centre. In: Proceeding of CHI 2008, Florence, Italy, April 5-10, pp. 1285–1294. ACM, New York (2008)
18. Ringle, C.M., Wende, S., Becker, J.-M.: SmartPLS-version 2.0. University at Hambur, Germany (2005), http://www.smartpls.de/
19. Schöning, J., Krüger, A., Olivier, P.: Multi-touch Is Dead, Long Live Multi-touch. In: Proc. of 27th Int'l Conf. Human Fact. Comput. Syst. CHI 2009, pp. 1–4 (2009)
20. Wang, Y.-S., Wang, H.-Y., Shee, D.Y.: Measuring E-Learning Systems Success in an Organizational Context: Scale Development and Validation. Computers in Human Behavior 23(4), 1792–1808 (2007)

Old Dogs Can Learn New Tricks: Exploring Effective Strategies to Facilitate Somatosensory Video Games for Institutionalized Older Veterans

I-Tsun Chiang

Graduate Institute of Sports and Health, National Changhua University of Education, No. 1, Jin-De Road, Changhua, Taiwan 50007
john@ncue.edu.tw

Abstract. The purpose of the study is to explore the possibility of utilizing somatosensory video games (SVGs) to promote institutionalized older veterans to participate regular exercises in senior institutions in Taiwan. Twenty-three veterans voluntarily completed 30-minute Wii Fit Plus for 3 time/wk for 8 weeks. Participation observation and informal conversation were two major qualitative research methods to collect responses and feedback of veterans who participants SVGs, co-workers, and facilitators. Constant comparison was utilized to analyze the qualitative information with QSR Nvivo7. The results identified that immediately feedback, competition, companionship, challenges, close to grandchild and fun are reasons to attract older veterans to keep involved with SVGs. The finding showed that SVGs are viable way to attract institutionalized older veterans to participate physical activities. The study implicates that enjoyment, social interaction and flow experience are keys to develop successful SVGs sessions for institutionalized older veterans.

Keywords: somatosensory video game, Wii Fit Plus, Elderly, leisure, recreation, health promotion, balance, technology, aging, aged, physical activity.

1 Introduction

More than 2 million older adults aged more than 65 years old reached 10.69% of total population in Taiwan in 2010 [1]. Among a variety of challenges in aged society, falls are the most significant serious problems and the leading cause of accidental death among older adults over 65 years of age [2,3]. In Taiwan, elderly fall is currently ranked as the second risk that caused death of elderly and the percentage of elderly fall has been increased from 18.7 % in 1999 to 20.5% in 2005 [4]. Personal and environmental factors are two major domains that cause elderly falls [5]. Personal factors are intrinsic causes that include balance impairment, neurological disorders, sensory deterioration, musculoskeletal disorders, medical uses and postural hypotension [2]. Poor lighting, ill-fitting footwear, slippery surfaces, wet floor and inappropriate furniture are some extrinsic reasons of environmental factors. Among those factors, impaired balance has been reported as one of the most important keys

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 88–100, 2012.
© Springer-Verlag Berlin Heidelberg 2012

among the multiple factors associated with falls in elderly due to visual, vestibular or somaesthetic sensory deficits, decrease in muscular strength, decrease in attention ability of performing multiple tasks, and other causes [6]. Therefore, postural control/stability interventions have been called to encounter the consequences of impaired balance in older adults. With the degradation of visual ability, musculoskeletal and somatosensory trainings are recommended to maintain the levels of balance in older adults [7].

To improve deficits in falls, fall prevention trainings are recommended to expose older adults to movement activities that mimic more skilled abilities of daily life such as twisting, turning, sudden starts and stops, standing over unstable surfaces, walking while changing speed and direction, using narrow paths, overcoming obstacles, and so on. In order to provide a variety of trainings, it regularly requires a certain amount to facilities and spaces to design. Therefore, those trainings are limited in research institutions and become inaccessible and unavailable for older institutions or communities. With the advance of technology, we proposed to develop a low cost, accessible, easy operate video game, Nintendo WiiFit ("WiiFit"), to proxy the high-cost, laboratory-based trainings and examine its applicability in senior institutions.

The WiiFit is a video game console of Nintendo® which designed to train balance, strength, flexibility and fitness [8]. The console includes a variety of software and a balance board (a 511 x 316 x 53.2 mm flat board) which incorporates pressure sensors to detect movements and monitor center of balance and shifts in player's body weight. The main software provides aerobics, balance, strength, and yoga programs which activities with increasing levels of challenges for the users to choose. Attractive designs, visual/audio feedback on accuracy, fun and motivating are reported as main characteristics to develop as a potential sustainable physical activity for elder adults [9]. Pilot studies identified that WiiFit is an acceptable fall interventions for elderly [8,9]. Sugarman, Weisel-Eichler, Burstin and Brown [10] also found that WiiFit is the potential to be used in clinical settings in order to improve balance in their feasibility study. However, there is still a lack of effective strategies to motivate and facilitate somatosensory video games (SVGs) for institutionalized older adults. This study aims to explore effective strategies and constraints of facilitating the SVGs for institutionalized veterans from qualitative perspective.

2 Research Methodology

In the veteran home, the study was advertised as training sessions for a "senior balance game" at the end of the year. A total of 631 veterans were recruited from a Veterans Home at Taoyuan County in Northern Taiwan. Participants received posters, fliers, staff persuasions, and mangers' announcement in the monthly meeting, individual invitations of oral introductions and video demonstration with PowerPoint presentation during 2-week recruitment.

2.1 Inclusion and Exclusion Criteria

Veterans aged 75 years or older, without physical disabilities and visual impairments that may interference the accessibility of playing the game, were recruited to participate.

Exclusion criteria were also executed to ensure the safety and to decrease the drop-off rate of participation. Individuals with wheelchair, having a long-term travel/leaving plan, intensive hospital care, or enrolling in other exercise/rehabilitation programs were excluded with euphemistic explanation.

2.2 Participants

Twenty-three of the participants (age: 82.4±3.9 years, height: 162.7±6.4 cm, weight: 67.4±6.7 kg) were willingly to be randomly assigned into two groups (yoga group and balance group) and join the 30-minute WiiFit sport SVGs for 3 time/wk for 8 weeks. All veterans participated in the study were pre-approved by the medical/managing staff and signed informed consent forms approved by the human subjects committee at the Taipei Physical College (Approval RefNo. 20090052, 12/3/2009) before participating in the study. In order to protect participants' privacy, pseudo names were used on all reported data.

2.3 SVG Programs

Yoga Group. Eleven participants (age: 82.9±3.9 years, height: 163.9±6.6 cm, weight: 65.8±5.2 kg)who were randomly assigned in the yoga group were allowed to choose trainings from 5 yoga poses which were included "Deep Breathing", "Warrior", "Half Moon", "Palm Tree" and "Chair" in the very beginning sessions (Figure 1). More yoga postural options were provided while the participants were familiar with the game and requested for more challenges. Weekly intervention plans were adjusted according to participants' average performance (Table 1).

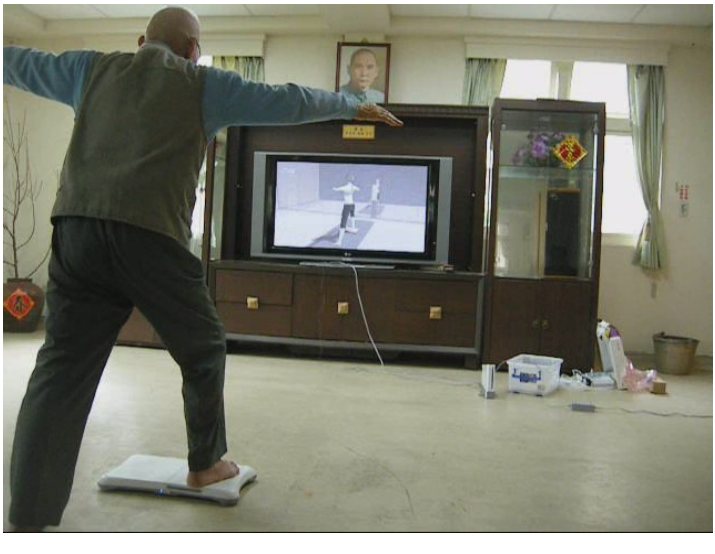


Fig. 1. One participant in Yoga group was playing "Warrior Position"

Balance Group. Twelve participants (age: 81.9±4.0 years, height: 161.6±6.2 cm, weight: 68.9±7.7 kg) who were randomly assigned in the balance group were allowed to choose trainings from 4 balance games which were included "Soccer Heading", "Table Tilt", "Balance Bubble", and "Penguin Slide" in beginning sessions (Figure 2). More game options were also provided when participants requested for more challenges (Table 1).

Fig. 2. One participant in Balance group was playing "Table Tilt"

All participants were trained one hour 3 times a week for 8 weeks. Within one hour training session, 5-minute warm-up, 50- minute game training, and 5- minute cool-down were provided in the leisure-activity room in veteran home. One primary investigator facilitated all sessions to ensure the quality and equality of interventions between groups and sessions. Participants with time conflicts (e.g. family member visiting, regular check-up, religious holidays…etc.) were allowed to adjust training blocks within the same weeks for maintaining 3 times a week. During the training, caring staffs were requested to monitoring their safety and physical performance and residential physicians and nurses were also on duty to ensure the access of emergency procedures.

2.4 Qualitative Data Collection and Analysis

Qualitative research methods, participation observation and informal conversation, were two major research methods to collect responses and feedback of elder adults who participants SVG programs, co-workers, and facilitators. Qualitative data were imported to a computer assisted qualitative data analysis software (CAQDAS), QSR*Nvivo7. Constant comparison was utilized to analyze the qualitative information [11].

Table 1. SVGs Training Plans

Time	Yoga Group (Game names x attempts)	Balance Group (Game names x attempts)
Week 1	Deep Breathing x 5 Half Moon x 6 Warrior x 5	Soccer Heading x12
Week 2	Deep Breathing x 3 Half Moon x 8 Warrior x 7	Soccer Heading x 12
Week 3	Deep Breathing x 3 Half Moon x 6 Warrior x 9	Soccer Heading x 10 Table Tilt x 5
Week 4	Deep Breathing x 3 Half Moon x 6 Warrior x 8	Soccer Heading x 6 Table Tilt x 12
Week 5	Deep Breathing x 3 Half Moon x 6 Warrior x 6	Soccer Heading x 6 Table Tilt x 12 Balance Bubble x 3
Week 6	Deep Breathing x 3 Half Moon x 6 Warrior x 6 Palm Tree x 5 Chair x 5	Soccer Heading x 3 Table Tilt x 5 Balance Bubble x 6 Penguin Slide x 4
Week 7	Deep Breathing x 3 Half Moon x 3 Warrior x 3 Palm Tree x 9 Chair x 5	Soccer Heading x 3 Table Tilt x 3 Balance Bubble x 12 Penguin Slide x 4
Week 8	Deep Breathing x 3 Half Moon x 3 Warrior x 3 Palm Tree x 10 Chair x 10	Soccer Heading x 3 Table Tilt x 3 Balance Bubble x 6 Penguin Slide x 9

3 Results

According to the analysis of qualitative information, the results showed that immediately feedback, competition, companionship, challenges, close to grandchild and fun were six main reasons to attract older veterans to keep involved with SVGs. Additionally, adding extra workload for staff, no interest in high-tech, game control/operation and amotivation for learning new things are four challenges to promote SVGs. In order to protect the confidentiality of participants, all participants were assigned pseudo names to use in the study.

3.1 Adherence with the Intervention

With flexible make-up training adjustment, all participants in both groups completed 24 sessions in the interventions (See Table 1). Seven sessions of 5 individuals were

adjusted and had make-up ones within the same weeks for personal reasons. The overall adherence with the intervention was 100 % and a total of 552 out of a possible 552 sessions (23 participants × 8 weeks × 3 sessions per week) were completed.

3.2 Qualitative Themes

Table 2 shows six themes (immediately feedback, competition, companionship, challenges, close to grandchild and fun) that were identified as main reasons to attract older veterans to keep involved with the intervention. In addition, enjoyment, social interaction and flow experience were three keys of interventions which could be found in those themes and kept older veterans involved with the game.

Table 2. Qualitative themes and factors of playing SVG programs for Veterans

Themes	Factors	Keys of the intervention
Immediately feedback	Feedback in Scores	Enjoyment
	Feedback from audience	Social interaction
	Audio and visual feedback	Flow experience
Competition	Score comparison with others	Enjoyment
	Compare their attendance of the intervention program	Social interaction Flow experience
Companionship	Encourage each other	Enjoyment
	Encourage the attendance	Social interaction
	Sharing the tips of getting higher scores	
	Caring about each other during the sessions	
Challenges	Feel satisfaction when passing stages	Enjoyment Social interaction
	Have strong desire to overcome next stages	Flow experience
	Want to keep practice	
Close to grandchild	Have common topics to share with grandchildren	Enjoyment Social interaction
	Can play it with grandchildren	
Fun	Laugh happily during the intervention	Enjoyment Flow experience
	Asking the extension of the time of training sessions	
	Arriving earlier than scheduled time	

Immediately Feedback. Immediate feedback is the first theme that could be observed during the training session. According to the nodes of field observation notes, the SVG provided immediately audio and visual feedback during the game. Players were easy to know if they did it right or wrong. Also, at the end of every session, the monitor also showed the scores of their performance. One participant shared his feedback with the primary investigator and said,

> *"Thank you so much for coming [and teaching us to play this video game]. Mr. Chang, it is so fun, exciting! It has beautiful pictures and lovely music without making too much noise. We never play such a game before. It is so high-tech! When I moved on the balance board, the character in the screen also moved. Really high-tech that we never catch up! When we were in military, we also did not have TV or computer. It was good enough to be able to survive. Thank you so much for sharing this with us!"*

Competition. Most of veterans still have the military spirit and enjoyed the competition. Veterans who participated the SVGs were able to chat each other during the training. They often discussed about their scores and compared with themselves and others. One participant told the instructor that,

> *"Teacher, did Arnold miss any sessions? I told him that he has to come to this training every time because it is good to his body and good for his emotion. I also told David that we are going to attend this training every time and get the full attendance award, ha ha ha! "*

Companionship. Companionship is another theme that generated from the field notes and interviews. They not only encouraged each other to attendance every session, but also encouraged their performance in the games by sharing tips of getting higher scores. In addition, they were caring each other for safety and health conditions during the intervention. Two participants shared their friends' physical conditions and said,

> *"Hi, Mr. Chang, Calvin is not going to come today because he has to see the doctor. He will come the day after tomorrow. Everyone loves to come to play this. Although Calvin's physical condition is not that good, but I still encourage him to come here and play this. It makes us feel happier together and forget our troubles."*

> *"Did Harry feel better now? It is so great to come here and do some exercise. We can chat each other, hang out together, make new friends. How wonderful"*

Challenges. There were lots of challenges during the game, such as the design of stages, difficulties, and scores. Those challenges made veterans are willing to keep practice and to overcome those challenges. For example, two participants discussed about their challenges,

> *"I discussed with Max yesterday and figured out the skills that I need to have to pass those stages. Both of us agree that we can perform better if we can have more time to practice."*

> *"Yes, I completed stage 3 yesterday! Max and Richard said that I am very good in that. But I think that stage 4 is very difficult for me. I never pass it."*

Close to grandchild. Because it was a very popular game among younger generations, this game make veterans have common conversation topics to share with their grandchild. In addition, they can also actually play with their grandchild after the training sessions. For instance, two participants shared their interactions with their grandchildren and said,

> *"Thank you so much for facilitating this activity. I had a family reunion last week and told them that I play Wii Fit recently. Both my son and grandson told me that I am so cool and catch up the trend. They encourage me to keep involving this activity and believe that this activity makes my life colorful here. They plan to buy one at home, and then, I can play with them when I come back home and see who is the winner."*

> *"I have practice this for a while and know how to play it. My grandchildren think that I am very skillful in playing this game in such an old age. Not only knowing, but also playing it so well. Grandchildren like to play with me every time when I go back home."*

Fun. It is easy to observe that veterans love this game. They always laughed happily during the game, asking to extend the time of playing and arrived earlier than scheduled time during the trainings. The following two observation notes can illustrate their enjoyment when they play this game.

> *"I observed that those veterans always come to the sessions before the scheduled time. You always can see the people for next time slot have arrived and are waiting there when I still have ten minutes to work with the people in previous time slot. For the people who are in the first time slot always come earlier and help me to set up the Wii Fit and related equipments, such as turning up the lights or turning on TV."*

> *"During the training, they always laugh loudly and said, "That is so fun!" Oliver, Paul, and Richard always ask if I can give them more time to play."*

3.3 Constraints

In order to understand the possibility to develop a sustainable project for using SVGs in older institutions, the study probed possible constraints in the study. Adding extra-workload for staff, no interest in developing technology-based intervention, difficulties in game control/operation and amotivation for learning are four challenges to promote somatosensory video game programs.

Adding extra-workload for staff. In order to facilitating this SVGs, the manger had to arrange a special room/space for the equipment storage and intervention. Also, before and during the intervention, primary investigator had to set up and operate this game during the intervention. In addition, staffs were requested to make up and clean out the room. All of those procedures would add extra-workload for staff. One manager expressed his opinion and said,

> *"As a governmental organization, we have very limited manpower in this facility. Every staff has his/her own duties already. For example, social workers, nurses, occupational and physical therapists may be suitable to ask them to do this program. However, they are not required to do this before. If I asked them to do it now, it is an extra-workload for them. They may be not happy with it…"*

No interest in developing technology-based intervention. In the real and diverse world, not everyone is interested in technology. Therefore, one staff stated that,

> *"We have some staff who are definitely not interested in those high technology at all. I don't think it will work if you ask them to develop this type of intervention…Every time we tried to encourage our older adults to try something new or to learn something related to technology, we had to learn first, but there must be some staffs jumped out and said no to it…"*

Difficulties in game control/operation. Because game bundle is not originally designed for older adults, the control or operation procedures are still too complex for elderly, even for some older staffs. The primary investigator recorded his observation notes when he operated this game and found its weakness,

> *"I observed that it is impossible for those veterans to operate this game by themselves. It is so complicated. It is not one button design. After turning it on, you have to use the Bluetooth remote controller to click here, click there, then here, then there…If it could be designed like TV, just turn it on and play, it will be much better. Therefore, I believe that user-friendly improvements should be one of the major considerations for the game producer."*

Amotivation for learning. Some veterans experience lots of frustrations and failures in their life and those life histories may cause their negative emotions and behaviors (e.g. learned helpless) in their daily life. One staff said,

> *"You never have a chance to get some people here. They have no motivation to everything, period…Really, they don't like this, they don't like that. It's very difficult to find something they like here. Maybe they had too much painful experience during the wars and in their lives. I am still figuring out how to help them."*

4 Discussion

With a variety of technology-based leisure interventions, a key reason to select the WiiFit programs of the study is because of the accessibility and popularity of the game. The Nintendo Fiscal year report stated that the Wii has sold 70.93 million units worldwide as of March 31 and 20.53 million units within 2009-2010 fiscal year [12]. According to the Entertainment Software Association (ESA) 2010 Essential facts, 26% of Americans over the age of 50 play video games, an increase from nine percent in 1999 [13]. In addition, SVGs have recently seen a growing trend in developing physical movement as a part of operation, such as Dance Dance Revolution (DDR) and ParaParaParadise by Konami. Based on gamers' dancing in time with a musical dance tune and moving graphical objects, those games promote physical activities positively (Designing physical and social intergenerational family entertainment). A research in Asia also found that technology-based video game with physical demands are attractive and exciting for older adults, especially when the games provide physical and social intergenerational family entertainment [14]. Reported as a fun, motivating, offering the potential for greater compliance, WiiFit is recognized as a potential solution to develop sustainability of physical activity programs [14]. In our study, participants also demonstrated high levels of interests, concentration, and cooperation resulting in increasing game performing, asking the extension of the time for training sessions and arriving earlier than the scheduled time during the intervention [15]. Thus, we believe that there is a great growing need and trend for SVGs which promote physical activity and mental stimulation for older adults.

4.1 Complexity on SVG Programs

According to the adherence with the intervention, qualitative themes, keys of the intervention and constraints, the findings can be explained and discussed by complexity theory. The theory has been pervasively used to explain clinical practice in a variety of professions. It emphasizes uncertainty and randomness constructs a non-linear dynamic system that traditional organizational theories are inability to explain and predict [16]. The theory which provides a graphical framework and creates a complex zone with appropriate degree of agreement and certainty between simple and chaotic situation and be used to illustrate our findings (Figure 3).

4.2 Strengths and Limitations

Mobily [17] reviewed the role of exercise and physical activity in therapeutic recreation services and found that leisure interventions are in the best position to deliver ongoing lifetime programs of regular physical activity that promote and maintain functional ability and health among persons with disabilities and chronic condition. A major strength of this study was to provide opportunities for institutionalized veterans to participate technology-based leisure/physical activities. Through the process of participating WiiFit, older veterans understood the popular trend of the society, rebuilt communication with communities, and expended their

spectrum of recreation and leisure awareness/appreciation. For example, some participants have learned that yoga and soccer are interesting in the WiiFit games and may join "real" yoga and soccer sessions when they are available in the veteran home.

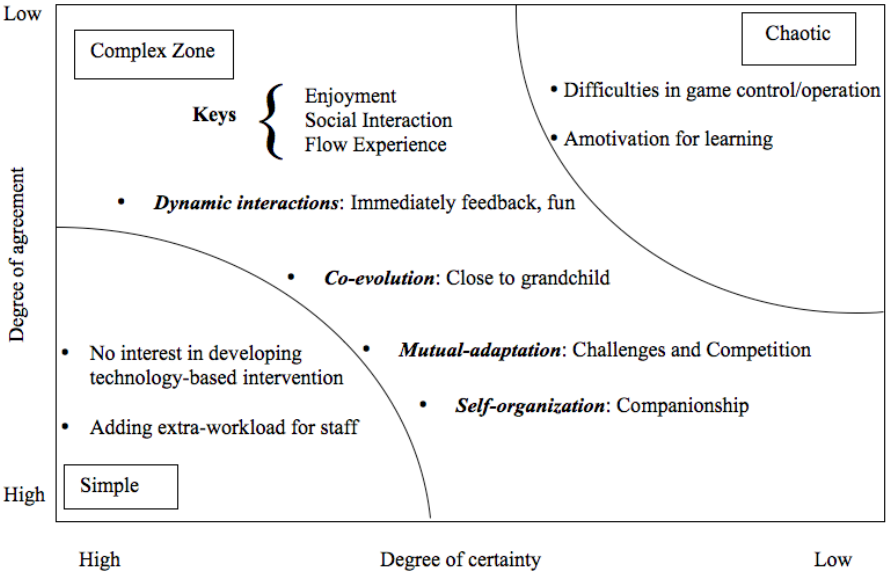


Fig. 3. Using complexity theory on SVG programs

The prevention of falls is currently one of the greatest challenges older adults experience in maintaining quality of life. Therefore, health- and fitness-related activities are important keys to enhancing the independence of older adults and are frequently offered in senior centers. The Second strength of the study is bridging sports, leisure and technology and provided evidence/outcome-based research for the leisure intervention. Without limitations of weather, space, physical limitation, or personnel, this intervention provided a highly attractive, interactive, and sustainable leisure activity for older veterans in their veteran home. Besides, the management team of the veteran home also has learned to explore and develop future possible alternative technology-based leisure interventions.

The study extends the empirical support that SVGs that focused on activity engagement and leisure enjoyment are viable approaches in senior institutions. Evidence that comprehensive and holistic fall-prevention studies of the response to WiiFit programs with other modes of fitness and balance trainings is recommended to expand the evidence base for this new approach. In addition, getting carried away with the game and fall, possibility of undesirable movements or postures, and overuse injuries are three major safety concerns of the study. Further safety guidance and information are also suggested in further studies.

5 Conclusions

An old age of idleness, without purposeful interventions to promote health and to help maintain identity, may undermine aging well in older adults [18]. The study gives insight on how leisure interventions were combined with technology and physical activity to enhance middle-old older veterans' leisure participation, improve their balance performance, and finally, promote their health. The research findings showed that SVG programs are viable way to attract institutionalized older veterans to participate initially. Further individualized programs and assistance in control and selections are needed to continue their involvement. The study implicates that enjoyment, social interaction and flow experience are keys to develop successful SVGs for institutionalized older veterans.

Acknowledgments. This study was supported by the Taoyuan Veterans Home, Veterans Affairs Commission and partially supported by the Department of Science Education, National Science Council, Executive Yuan, Taiwan under Grant No. NSC 99-2511-S-018-019. In addition, thank to Mr. Pin-Shih Chang for his assistance in implementation and data collection.

References

1. Ministry of the Interior, Taiwan. Interior National Indicators.: Percentage of Population for Ages 65 and Over by Selected Counties, The Ministry of the Interior (2010)
2. Hornbrook, M.C., Stevens, V.J., Wingfield, D.J., Hollis, J.F., Greenlick, M.R., Ory, M.G.: Preventing falls among community-dwelling older persons: results from a randomized trial. Gerontologist 34(1), 16–23 (1994)
3. Nelson, R.C., Amin, M.A.: Falls in the elderly. Emergency Medicine Clinics of North America 8, 309–324 (1990)
4. National Health Research Institutes, http://enews.nhri.org.tw/enews_list_new2.php?volume_indx=182&showx=showarticle&article_indx=5503&enews_dt=2006-12-28
5. Ashley, M.J., Gryfe, C.I., Amies, A.: A longitudinal study of falls in an elderly population, II: some circumstances of falling. Age Ageing 6, 211–220 (1977)
6. Muir, S.W., Berg, K., Chesworth, B., Klar, N., Speechley, M.: Balance impairment as a risk factor for falls in community-dwelling older adults who are high functioning: a prospective study. Physical Therapy 90(3), 338–347 (2010)
7. Prado, J.M., Stoffregen, T.A., Duarte, M.: Postural sway during dual tasks in young and elderly adults. Gerontology 53(5), 274–281 (2007)
8. Nitz, J.C., Kuys, S., Isles, R., Fu, S.: Is the Wii Fit a new-generation tool for improving balance, health and well-being? A pilot study. Climacteric 13(5), 487–491 (2010)
9. Williams, M.A., Soiza, R.L., Jenkinson, A.M., Stewart, A.: EXercising with Computers in Later Life (EXCELL) - pilot and feasibility study of the acceptability of the Nintendo® WiiFit in community-dwelling fallers. BMC Research Notes 3, 238 (2010)
10. Sugarman, H., Burstin, A., Weisel-Eichler, A., Brown, R.: Use of the Wii Fit System for the treatment of balance problems in the elderly: A feasibility study. In: Virtual Rehabilitation International Conference, Israel (2009)

11. Strauss, A., Corbin, J.: Basics of qualitative research: Grounded theory procedures and techniques. Sage, London (1990)
12. Nintendo Fiscal year report, http://www.gamespot.com/news/6261400.html?tag=recent_news
13. Entertainment Software Association, http://www.theesa.com/facts/
14. Hughes, S.L., Seymour, R.B., Campbell, R.T., Whitelaw, N., Bazzarre, T.: Best-practice physical activity programs for older adults: findings from the National Impact Study. American Journal of Public Health 99, 362–368 (2009)
15. Andrews, A.W.: Feasibility and benefit of using the Nintendo Wii fit for balance rehabilitation in an elderly patient experiencing recurrent falls. J. Student Physical Therapy Research 1(2), 12–20 (2010)
16. Grobman, G.M.: Complexity theory: A new way to look at organizational change. Public Administration Quarterly 29(3), 351–384 (2005)
17. Mobily, K.E.: Role of exercise and physical activity in therapeutic recreation services. Therapeutic Recreation Journal 43(2), 9–26 (2009)
18. Hawkins, B.A.: Aging well: toward a way of life for all people. Preventing Chronic Disease 2(3), A03 (2005)

Effects of Multi-symbols on Enhancing Virtual Reality Based Collaborative Task

Shih-Ching Yeh[1], Wu-Yuin Hwang[2,*], Jing-Liang Wang[1], and Yuin-Ren Chen[1]

[1] Department of Computer Science and Information, National Central University, Taiwan
[2] Institute of Network Learning Technology, National Central University, Taiwan
wyhwang1206@gmail.com

Abstract. Applying virtual reality (VR) technologies to enhance learning becomes more and more popular. This research intends to investigate how multi-symbolic representations could help users being aware of collaborative context and partner's needs to enhance completing haptics-based collaborative tasks in a co-located/distant virtual environment. This study evaluates the performance of collaboration including the completing time and the number of failure in completing a task. To make users being aware of context, multi-symbolic representations in forms of color and text are provided as well as haptics and audio feedback in the virtual environment. Participants in the experiment were separated into four groups with the combinations of two variables: w/o multi-symbols and co-located/distant. The results show that multi-symbols significantly helped users reduce the time in completing a task in the case of co-located collaborative virtual environment. However, there was no significant improvement in performance in the case of distant collaborative virtual environment. From our on-site observation, we found users had less verbal communication in strategies to complete the task though microphone and speaker provided on both sides within a distant virtual environment, therefore, effects of multi-symbols cannot be revealed to the performance of collaboration. Furthermore, to investigate how multi-symbols could affect user's perceptions, we investigate the perceived awareness, presence and social presence of our proposed system and its influence on perceived usefulness, ease of use and playfulness based on Technology Acceptance Model. The results showed that awareness, presence and social presence significantly influenced perceived usefulness, ease of use and playfulness. Therefore, our proposed multi-symbols virtual reality system has potentials to help collaborative learning.

Keywords: Virtual reality, haptics, collaboration, multi-symbols.

1 Introduction

With the fast growing of computer and internet technologies, digital learning has been widely applied in many aspects and many applications for learning are developed and

* Corresponding author.

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 101–111, 2012.
© Springer-Verlag Berlin Heidelberg 2012

keeping explored. In contrast to conventional learning in the classroom, digital learning enables learners to participate the learning activity and interact with digital learning platform with more flexibility, for example, a adaptive learning schedule or an flexible learning time…etc. Furthermore, a digital learning system is able to adaptively plan a learning schedule for the learner based on learning portfolio and needs with an intelligent system. Though digital learning can offer many advantages, there is still a lot unknowns existing in their effects to learning. As a result, many researchers have been focusing on the study of learning technologies or activities and investigate their learning effect and its related issues, such as the Virtual Reality (VR) or augmented reality and the evaluation of their learning effect or the improvement of learning effect. To build an immersive learning environment for learners to be engaged or interact with, VR technologies are often proposed and applied. For example, Konstantinidis et al. utilized Second Life to construct a collaborative learning platform [8], Tina et al. integrated Wonderland with various sensory feedbacks, including stereo imaging, spatial audio, haptics to build a virtual collaborative learning platform that students were able to work together with teachers. [19].

Virtual Reality (VR) is to simulate a virtual environment in which users are able to interact with and being engaged via various sensory feedbacks, such as stereo imaging, spatial audio, haptics or smell. It has been widely applied to many fields like medical rehabilitation, military training, digital learning…etc. Further, researchers have been trying to create collaborative tasks within virtual environments therefore to improve learning. Besides, symbolic representations are also considered as important factors that are able to facilitate learning. Using symbols to represent phenomenon of math, physics or sociology can advance the comprehension of learning contents from learner perspective therefore to build the links among the problem itself, its solving process and the real-world tasks [12].

This research intends to investigate how multiple symbolic representations could help users being aware of context and enhance the performance of collaboration in completing a haptics-based task within a co-located/distant virtual environment. Also, interaction and communication between members are observed in both cases (co-located/distant). Further, to investigate user's perceptions while multi-symbols are provided, we investigate the awareness, presence and social presence of our proposed system and its correlation with usefulness, ease of use and playfulness based on Technology Acceptance Model.

2 Checking the PDF File

2.1 Representation, Awareness and Presence

Representations are the fundamental products of a series of cognition process. We are able to deeply understand or realize the connotation of knowledge via various forms of representation. In math education, [9] proposed five types of representations: object representation, concrete representation, arithmetic symbol representation,

spoken-language representation and picture or graphic representation. The utilization of these types of representation can make the process of problem-solving more natural and more reasonable. In application of augmented reality, [14] stated that representation in the form of shadow can make the virtual object more realistic and more stereo. In this research, multi-symbolic representations are applied in a VR-based collaborative task to help users being aware of the context, for example the desire of partner, the status of the collaborative tasks.etc., therefore to make appropriate operation, communication and coordination in order to enhance the performance in completing the collaborative task.

While working with a collaborative task, group awareness is considered as a crucial factor to complete the task effectively. Daniel et al.[5] indicated that three types of group awareness were proposed and studied: behavioral awareness, cognitive awareness and social awareness; they are considered as crucial for effective collaborative learning. Ferran et al.[6] also concluded that mutual awareness is highly decreased therefore to hinder direct communication [6].

In virtual environment, presence stands for the extent of immersion or engagement between users and virtual environment. [16] mentioned that factors affecting engagement include ease of interaction, user-initiate control, pictorial quality, length of exposure and social factors, which are related to presence Among these factors, interaction and control were considered as the most important. The next question is how to define and how to measure the presence. Bob G. and Michael J(1998) proposed a questionnaire to measure presence[2].

As a result, this research utilized several questionnaires based on the study of Bob and Michael (1998) to measure the awareness and presence from user perspective and investigate their correlations with the performance in completing the task in a collaborative virtual environment.

2.2 Social Presence and Collaboration in Virtual Reality

Short and Christie (1976) proposed the concept of social presence which means the extent a media may bring remote users closely [18]. This is trying to explain the functional role a media may play during the process remote users interact with each others and further describe the interaction behavior existing in communication media [18][22]. However, the lack of verbal or non-verbal cues may reduce the social presence [21]. Newberry (2001) indicated that the raise of social presence in online environment may help students create more deep impressions on that related experience [13]. Rourke et al.(1999) concluded that all participants may perceive warm, collegial, and approachable with learning environment having high levels of social presence. High levels of social presence can be achieved by making group interactions appealing, engaging, and intrinsically rewarding. This can further instigate, sustain, and support cognitive and affective learning objectives[17].

Collaborative learning played an important role in distant education and was widely applied while there have been many researches to investigate its importance. Susan mentioned that students in distant education may create better outcomes while

opinion sharing and discussion were provided via collaborative learning [15]. Lee et al.(2010) indicated that, in a VR-based learning environment, interaction experience and learning experience can enhance learning effect while learning experience is determined from several factors, such as presence, motivation, cognitive benefits, control and active learning.[1].

Symbol is the fundamental product of a series of cognition process and is one of the key factors that may determine learning effect. While two users are side-by-side in the co-located environment and work collaboratively, they can have various forms of instant communication via the revelation of multiple representations like audio or gesture and plan for the strategy to complete the task. However, while putting two users at distant location with the support of audio communication as well as the revelation of multiple symbolic representations, are multi-symbols still helpful for effective learning? This research intends to investigate if there is any difference in performance while operating collaborative task co-located/distant.

3 Method

This study aims at the effect of multi-symbols helping with the improvement of performance in completing a collaborative task within virtual environment. The improvement of performance is indexed with the decrease of failed number and completing time. Further, this research looks into the difference of the effect of multi-symbols operating at co-located environment and distant environment. Also, questionnaires were designed to measure presence and awareness from user perspective. Correlations between multi-symbols and user perceptions are studied.

Our hypotheses are listed below:

H1: In the case of distant collaboration, multi-symbols can help to reduce failed number.

H2: In the case of distant collaboration, multi-symbols can help to reduce the completing time.

H3: In the case of co-located collaboration, multi-symbols can help to reduce failed number.

H4: In the case of co-located collaboration, multi-symbols can help to reduce the completing time.

H5: Without the help of multi-symbols, participants in the case of co-located collaborative work can have a better performance in the time to complete the task.

H6: Without the help of multi-symbols, participants in the case of co-located collaborative work can have a better performance in the failed number.

H7: With the help of multi-symbols, participants in the case of co-located collaborative work can have a better performance in the time to complete the task.

H8: With the help of multi-symbols, participants in the case of co-located collaborative work can have a better performance in the failed number.

3.1 Collaborative Task Design

In the virtual environment, there was a virtual cube appearing at different location and with varied size and weight. In each group, there are two participants; each participant handled a 3D haptics controller in order to control a virtual paddle. Two participants have to work collaboratively therefore to pinch the virtual cube, with two virtual paddles from left side and right side, and to lift the virtual cube to collide with a cone-like target at the central-top of the scene, shown in Fig. 2. The time limit for each trial was set 60 seconds. Each group had to successfully complete 10 trials.

The main difficulty of the task is two participants had to work coordinately, including the force output (pinch and lift) and moving speed, in order to complete the task.

Symbolic Representation Design. To make users being aware of the context, multi-symbols were designed in forms of color and text to provide users with appropriate instant messages as clues in order that users can complete the task effectively (see Fig. 2). First, to indicate the "touch" of virtual paddle with virtual cube, the color of virtual paddle changed upon the touch happened (left: from light blue to dark blue; right; left: from light red to dark red). Second, the color of virtual cube changed from light blue to light green while both paddles touched the virtual cube. It was a symbolic message to both users that the cube is ready to be lift. Third, text indicated the weight of the virtual cube was shown on the surface of cube. It was to make users being aware of the force they might have to enforce. Fourth, forces from both sides exerted on the cube were displayed in order that users were able to maintain the balance and work coordinately.

(a)Shutter glasses (b)Haptics device

Fig. 1. Sensory feedback

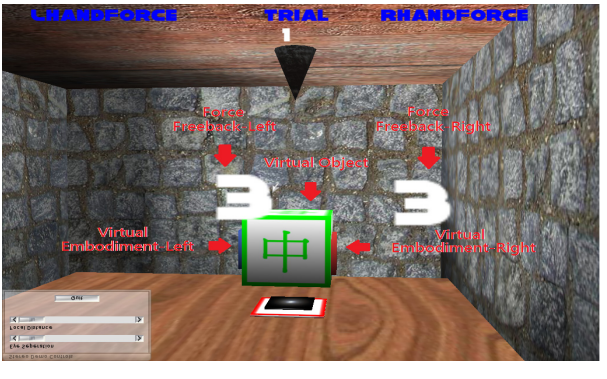


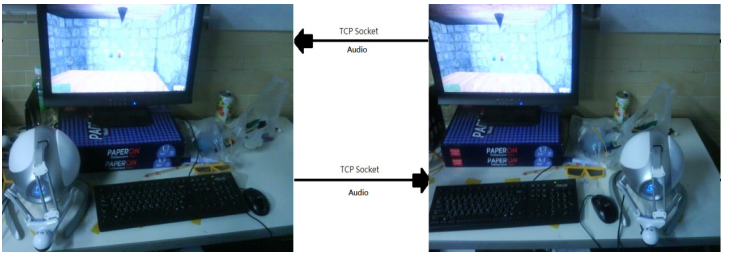
Fig. 2**.** Representation of multi-symbols

Co-located/distant design and setup. The setup of the experiment was classified into two different modes: co-located and distant, in order to investigate the effect of multi-symbol in both modes.

Co-located: Participants were put side-by-side in front of a large stereo display and worked coordinately via 3D haptics controller to complete the task. Both participants were free to communicate with each other, described in Fig. 3(a).

Distant: Participants were put at remote locations with his/her own 3D haptics controller and display. The scene appearing on two displays were synchronized. Also, microphone and speaker were provided on both sides that both participants were able to have verbal communication, shown in Fig. 3(b).

(a) Co-located case



(b) Distant case

Fig. 3. System setup

3.2 Participants and Experiment Procedures

Total 60 participants were recruited for the experiment and separated into 30 groups with two participants each group. These participants were all college students with computer related majors.

Considering multi-symbols and the co-located/distant setup, four test cases were designed as below:

Case1: Distant collaborative work with multi-symbols provided.
Case2: Distant collaborative work without multi-symbols provided.

Case3: Co-located collaborative work with multi-symbols provided.
Case4: Co-located collaborative work without multi-symbols provided.

Thirty groups were further classified into 4 clusters with each going through four cases in different order therefore to eliminate the effect of experimental sequence and their side effect. The experimental order assigned to each cluster is listed as below:

Order 1: Case1 -> Case 2 -> Case 3 ->Case 4;
Order 2: Case3 -> Case 4 -> Case 1 ->Case 2;
Order 3: Case2 -> Case 1 -> Case 3 ->Case 4;
Order 4: Case4 -> Case 3 -> Case 2 ->Case 1.

Questionnaires were filled at the end of the experiment.

Each participant's performance and perceptions were measured through the collaborative work in completing the task. Two performance indexes, the "failed number" and the "time" in completing a task, were defined and measured. Also, the time history of 3D position of each virtual paddle was measured and recorded at each time interval for further behavioral analysis.

4 Results and Discussions

To compare the performance of Case 1 and Case 2, independent samples T test was made to performance indexes (failed number & completing time). Results (see Table 1) show that the difference in completing time is not significant ($F(1,58)=1.216$, $p>0.05$) as well as failed number ($F(1,58)=0.493$, $p>0.05$)). As a result, H1 &H2 are false. Namely, in the distant case, multi-symbols do not help the collaborative work being complete more effectively.

To compare the performance of Case 3 and Case 4, independent samples T test was made to performance indexes (failed number & completing time). Results (see Table 2) show that the difference in completing time is significant ($F(1,58)=2.112$, $p<0.05$). Instead, results of failed number are not significant ($F(1,58)=4.611$, $p>0.05$). As a result, H3 is true and H4 is false. Namely, in the co-located case, multi-symbols do help the collaborative work being complete more effectively but not in failed number.

According to results shown above, we found that multi-symbols effectively enhance the performance in co-located case but not in distant case. From our on-site observation, in the distant case, participants did not communicate actively and efficiently though microphone and speaker provided. It is one of the causes that the effect of multi-symbols is not revealed.

To compare the performance of Case 2 and Case 4, independent samples T test was made to performance indexes (failed number & completing time). Results (see Table 3) show that the difference in completing time is not significant($F(1,58)=1.249$, $p>0.05$) as well as failed number ($F(1,58)=0.14$, $p>0.05$). As a result, H5 &H6 are false. Namely, without the support of multi-symbols, co-located case and distant case have similar performance while operating collaborative work.

To compare the performance of Case 1 and Case 3, independent samples T test was made to performance indexes (failed number & completing time). Results (see Table 4) show that the difference in completing time is significant($F(1,58)=2.090$, $p<0.05$) as well as failed number ($F(1,58)=7.387$, $p<0.05$). As a result, H7 &H8 are both true. Namely, with the support of multi-symbols, co-located case has a better performance than distant case while operating collaborative work.

According to results shown above, we found that, with the support of multi-symbols, co-located case had a better performance (both failed number and completing time) than distant case. However, without the support of multi-symbols, the performance of co-located case and distant case were close (difference is not significant). We believe that in distant case, each participant is not visible to the other that there is the lack of information regarding partner's body gesture and their needs therefore not able to understand partner's desire or action. Also, the willingness to interact with partner is reduced while feeling the partner is far away in the distant case. Thus, even with the support of multi-symbols, the performance is not improved.

While there is no multi-symbols provided, the information exchanged or shared between two participants is not easy to make any further communication or coordination to efficiently complete the collaborative task only via the haptics feedback or audio communication It showed that multi-symbols for collaborative task play vital roles to help collaborative task. From our on-site observations, while operating the task collaboratively, several interesting behaviors existing between participants, such as strategy toward task success or failure, were found. First, after a few trials of completing the task in the beginning, instead of pinching and lifting the virtual cube directly toward the cone-like target, participants first push the cube and slide it to the underneath of the cone-like target, and then lifted it upward till reaching the target. Namely, participants were able to develop a good strategy with less completing time or errors therefore to complete the task more efficiently and successfully even though it was in a virtual environment. Second, the failure in completing the task was generally caused by inconsistent and incoordinate movements or force between partners. While intending to pinch the virtual cube, a balanced force output from left side and right side is required to prevent the sliding. To lift the virtual cube, a coordinate movement in 3D space was important as well, otherwise, the virtual cube might easily fall down to the ground. Thus, the lack of communication or message exchange was one of the main reasons to cause incoordinate movements or force output and finally led to the failure in completing the task.

While results showed that the support of multi-symbols in the co-located case did not reduce the failed number effectively. From our on-site observation, we found that discussion, sharing and communication between group members could be important factors that might lead to the planning of a winning strategy. Though multi-symbols did provide participants more collaborative status information in forms of color and text, it didn't directly contribute to increase the frequency of discussion, share or communication between group members. A similar situation existed in the distant case. Though microphone and speaker were provided in the distant case, we noticed there was less interaction between members than in co-located case so that the effect

of multi-symbols were not revealed to performance indexes (failed number & completing time) in the distant case. According to the interview of each participant, it was found to have less sufficient interaction between group members, particularly in the distant case. The most common opinions in interviewing students were "My partner didn't move the way I desired" and "My partner always ignored my request or didn't want to communicate with me". In the distant case, opinion like "My partner pushed too much to react from my side" is a lot as well, despite that microphone and speaker were provided. In conclusion, except providing participants with multi-symbols, other interfaces or media, besides microphone and speaker, are needed to enable a higher frequency and efficiency of discussion, share and communication between group members therefore to improve the performance in a collaborative work.

5 Participants' Perceptions toward VR-Based Collaboration

Among many researches in the past, ease of use and usefulness are considered as the most important factors to determine if an user will accept a new type of technology system or not [7][10]. Davis (1989) defined usefulness as "the degree of which a person believes that using a particular system would enhance his or her job performance" and ease of use as "the degree of which a person believes that using a particular system would be free of effort" [4]. Vasileios and Anastasios (2010) also pointed out that the behavioral desire of computer based assessment is significantly correlated with playfulness and ease of use [20]. In this regard, ease of use, usefulness and playfulness are the three key elements to determine if an user will accept a new type of technology system. Other than the above mentioned factors, we consider that there are more factors may contribute to usefulness and playfulness. As a result, we propose three more items considered: presence, awareness and social presence. This research utilizes questionnaires to measure items above and investigates their correlations. We intend to investigate the correlations between (awareness, presence, social presence) and (ease of use, usefulness, playfulness).

While conducting a collaborative task in virtual environment, users need to be aware of context, such as task contents, partner or self, in order to have effective communication and coordination that lead to a success in completing the task. Thus, we intend to utilize multi-symbols and merge them with the virtual environment in the purpose to keep users being informed with instant status of the environment. In the virtual environment, Mania and Chalmers (2000) mentioned that presence is positive correlated with the learning effect [11]. Moor indicated that social presence affects not only learning effect but also the satisfaction of learning materials. We believe a collaborative VR task with better feedback in presence and social presence from user perceptions can contribute to effective performance in completing the task. Questionnaires are designed to measure presence and social presence.

After analyzing the questionnaires results, it was found, in either case (co-located/distant), the raise of users' awareness, presence and social presence may

enhance users' perceptions of usefulness, ease of use and playfulness therefore to foster the acceptance to new type of technology system.

6 Conclusion

The results show that multi-symbols significantly helped users reduce the time in completing a task in the case of co-located collaborative virtual environment. However, there was no significant improvement in performance in the case of distant collaborative virtual environment. From our on-site observation, we found users had less verbal communication, discussion or share in planning strategies to complete the task in the distant case, despite that microphone and speaker were provided. As a result, effects of multi-symbols can't be revealed to the performance in that case. In conclusion, except providing participants with multi-symbols, alternate interfaces, other than microphone and speaker, are needed to enable a higher frequency of discussion, share and communication between team members therefore to improve the performance in a collaborative work. Further, regarding the perceptions toward our proposed system, ANOVA test is made to each item. The results show that awareness, presence and social presence are highly correlated to usefulness, ease of use and playfulness. The raise of users' awareness, presence and social presence may enhance users' perceptions of usefulness, ease of use and playfulness therefore to foster the acceptance to new type of technology system.

References

1. Lee, E.A.-L., Wong, K.W., Fung, C.C.: How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach (2010)
2. Witmer, B.G., Singer, M.J.: (1998)
3. Burdra, G., Coiffet, P.: Virtual Reality Technology (2003)
4. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology (1989)
5. Suthers, D.D., Hundhausen, C.D., Girardeau, L.E.: Comparing the roles of representations in face-to-face and online computer supported collaborative learning. Laboratory for Interactive Learning Technologies (2003)
6. Argelaguet, F., Kulik, A., Kunert, A., Andujar, C., Froehlich, B.: See-through techniques for referential awareness in collaborative virtual reality (2011)
7. Keil, M., Beranek, P.M., Konsynski, B.R.: Usefulness and ease of use: field study evidence regarding task considerations (1995)
8. Andreas, K., Tsiatsos, T., Terzidou, T., Pomportsis, A.: Fostering collaborative learning in Second Life: Metaphors and affordances (2010)
9. Lesh, R., Post, T., Behr, M.: Representations and translations among representations in mathematics learning and problem solving. In: Janvier, C. (ed.) Problems of Representation in the Teaching and Learning of Mathematics, pp. 33–40. Lawrence Erlbaum, Hillsdale (1987)
10. Malhotra, Y., Galletta, D.F.: Extending the technology acceptance model to account for social influence: theoretical bases and empirical validation (1999)

11. Mania, K., Chalmers, A.: A user-centered methodology for investigating presence and task performance (2000)
12. National Council of Teachers of Mathematics. Principles and standards for school mathematics, Reston, VA (2000)
13. Newberry, B.: Raising Student Social Presence in Online Classes, In: WebNet 2001, Proceedings of the World Conference on the WWW and Internet (2001)
14. Sugano, N., Kato, H., Tachibana, K.: The Effects of Shadow Representation of Virtual Objects in Augmented Reality (2003)
15. Malmo, S.: Collaboration at a Distance:Improving Collaborative Efforts in a Distance-Education Environment (1999)
16. Stanney, K.M., Salvendy, G., Deisigner, J., Dizio, P., Ellis, S., Ellison, E., et al.: Aftereffects and sense of presence in virtual environments: Formulation of a research and development agenda (1998)
17. Rourke, L., Anderson, T., Garrison, D.R., Archer, W.: Assessing Social Presence in Asynchronous Text-Based Computer Conferencing (1999)
18. Short, J., Williams, E., Christie, B.: The social psychology of telecommunications. Wiley, London (1976)
19. Scheucher, T., Bailey, P.H., Gütl, C., Judson Harward, V.: iJOE: Collaborative Virtual 3D Environment for Internet-accessible Physics Experiments. In: REV 2009, vol. 5 (special issue 1), pp. 65–71 (August 2009) (republished)
20. Terzis, V., Economides, A.A.: The acceptance and use of computer based assessment (2011)
21. Williams, F., Rice, R.E.: Cimmunication research and the new media technologies (1983)
22. Westmyer, S.A., DiCioccio, R.L., Rubin, R.B.: Appropriateness and effectiveness of communication channels in competent interpersonal communication. Acceptance model for social influence: theoretical bases and empirical validation (1998)

Particle Filter with Affine Transformation for Multiple Key Points Tracking

Sheng Liu[1,2], Ting Fang[1], Shengyong Chen[1,⋆], Hanyang Tong[1], Changchun Yuan[1], and Zichen Chen[2]

[1] Zhejiang University of Technology
Hangzhou, 310023, Zhejiang, P.R. China
[2] Zhejiang University
Hangzhou, 310027, Zhejiang, P.R. China
edliu@zjut.edu.cn

Abstract. This paper proposes an accurate method for multiple key points tracking in long microscopic sequences. Tracking in normal-scale image sequences is proved to be a valuable fundamental technology in computer vision, while tracking in microscopic sequences is a more challenging work due to its poor image quality resulted from the complexity of microscopic imaging process. The micro stereo imaging process can be implemented in a tilting rotation of the stage which produces an affine geometric transformation on the projection of rigid spatial micro structure. This paper finds that the projection's affine invariance leads tracking of point templates to be a feasible solution, due to the fixed spatial relationship among the composed of simple fundamental components such as points, lines and planes. At the same time, we apply an adaptive particle filter (PF) of points tracking algorithm to sample and calculate the weights from those multiple point templates, which can resolve the visual distortion, illumination variability and irregular motion estimation. The experimental results are precise and robust for rigid multiple key points tracking in long micro image sequences.

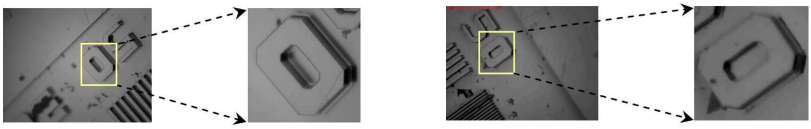
Keywords: Multiple points tracking, rotational microscopic vision, template matching constraint, particle filter with affine transformation.

1 Introduction

Tracking of feature points in microscopic videos is driven by potential applications such as medical virus and cells studies [1], surveillance systems [2], vehicle navigation systems [3], human-computer interaction [4], et al. It also plays an important role in many computer vision applications as image 3D reconstruction [5], segmentation [6], and recognition [7]. But there is little research focused

⋆ This work was supported by the National Natural Science Foundation of China (NSFC-60573123,60605013,60870002, 60802087), NCET, and the Science and Technology Dept of Zhejiang Province (2009C21008, 2010R10006, Y1090592,Y1110688).

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 112–126, 2012.
© Springer-Verlag Berlin Heidelberg 2012



(a) Normal microscopic image (b) Motion blur microscopic image

Fig. 1. The quality of images from rotational sequences are significantly worse than that of the still microscopic images, unclear contours are produced by the vibration of movements. In each sub image, from left to right are original image and regional amplified details respectively.

on the micro-scale applications except for medical applications. It's well known that the needs of 3D information acquirement and high-precision error detection of micro objects are highly required. Because the correspondences of feature points in image sequences are the bases of 3D structure reconstruction, feature points tracking among micro image sequences are necessary and valuable. Existing feature point tracking algorithms are usually classified as based on template, motion parameters, and color patch [8]. However, occlusions, image noise, and repeated texture remain challenging problems for multiple key points tracking in image sequences, especially for the low quality microscopic image sequences with weak texture. The quality of microscopic video is affected by motion blur (Fig.1) caused by the rotational motion seriously, as well as illumination condition, image noise and specular reflection.

These difficulties often lead to track failures for many state-of-the-arts tracking algorithms. The representative research results include:

For those image sequences with rich color information. Bradski presented the Continuously Adaptive Mean Shift (Camshift) algorithm [9], which takes color histogram as object mode for tracking in rich color images. But Camshift did not perform well in tracking the feature points on the condition of complicated background with areas of similar color.

For feature point tracking with given motion parameters. Yi-Sheng Yao et al, first designed a probabilistic data association filter combined with an Extended Kalman Filter (EKF) [10] to estimate the rotational motion and continuously track the feature points in frames, which effectively resolved occlusion problem. Nevertheless, the real location must be predicted by probability analysis during object arbitrary moving process, this makes probabilistic data association filter difficult to build precise motion model.

For non-rigid tracking. Aeron Buchanan et al. proposed a combining local and global motion models with Kanade-Lucas-Tomasi (KLT) tracker to accurately track multiple feature points of non-rigid moving objects [11].However, if motion predictions can not be made for the sub-sequences consisted of the first initial frames, this strategy must be failed.

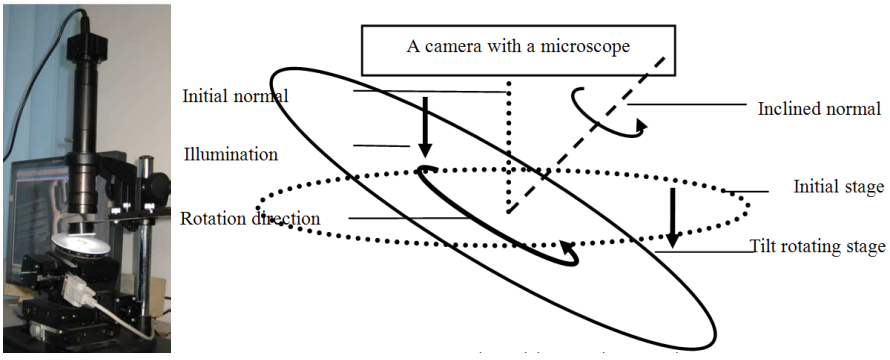


Fig. 2. Monocular microscopic vision system and rotational motion model

2 Related Work

The most relative work to ours was presented by Kwon et al. in 2009. Their method is a template tracking approach based on applying the particle filtering (PF) on the affine group, which can accurately track a large single rotational object region [12]. But it only performs well in the single template tracking, therefore we extended their method to the multiple points tracking case in this paper. Theoretically we know that the motion of micro objects with a given speed on the stage can be a parametric model, so does their microscopic projective images. But the tiny mechanical errors resulting from rotation movements of the stage will induce the movement of the micro object fail to follow the expecting motion model significantly. It means that both rotation transformation and shape deformation are existed in our case. As described in reference [12], when a highly nonlinear measurement function with respect to the state and a curved space state occurred, it is difficult to find a systematic way to choose tuning parameters for unscented transform (UT). So Unscented Particle Filter (UPF) is not suitable for feature points tracking in our case, that's why we chose the PF algorithm for the microscopic image tracking.

We design a monocular microscopic vision system for obtaining the 3D information of micro objects, as shown in Fig. 2. It mainly consists of the microscopic imaging device and visual information processing system. A tilting rotation movement is needed for capturing the video (stereo image sequences) with 3D depth information The micro object executes an inclined rotating movement by using the 4-DOF stage during the measuring process, while the microscopic imaging system is fixed at a perpendicular orientation. The advantages of sharing the constant intrinsic parameters by using single camera will ensure lower 3D reconstruction errors. Within an appropriate depth of field, the more the tilting angle is, the clearer projection image of micro 3D object details would be obtained. The deviation between the rotation axis and the field of view of microscope easily leads to the measuring micro object roll out of the view. To

acquire a long stereo image sequence, this deviation must be adjusted as small as possible before the measurement start.

Our proposed method successively deal with above mentioned problems and achieved accurate multi-points tracking in microscopic image sequence. As the tilt stage and rotational motion lead to both of the state space and its subspace are nonlinear, PF tracking algorithm with affine transformation is adopted to track the multiple points in long microscopic sequences. Since the images of the rigid measuring micro object have a fixed spatial relationship of global structures and local features affine invariabilities during the rotating process, the tracking problem is greatly simplified by combining with affine transformation and covariance descriptor (see Section 3.2 for the details).

This paper is organized as follows. Section 3 describes the theory of PF algorithm combined with affine transformation for multiple points tracking. Section 4 presents the experimental tracking results by using proposed PF and KLT algorithms, both on single point and multiple feature points in a microscopic image sequence. Then the errors are analyzed. In Section 5, conclusions are drawn.

3 Multiple Points Tracking Theory

The goal of multiple feature points tracking is to follow those given points and wish to output their accurate locations in a time dependent microscopic image sequence. Due to the weak texture, noise and motion blur, the details of a single point $b_t^{(i)}$ cannot be described accurately, so this paper takes the advantage of template patch to describe and track in a time-dependent micro image sequence. Giving $b_t^{(i)} = (x_t^{(i)} - y_t^{(i)})^T$, a column vector containing the pixel coordinates. Here $(i = 1, \cdots, k)$ indicates the number of tracked points in each frame; and $(t = 1, \cdots, N)$ indicates the number of frames. For multiple points initializing with giving a set of points center coordinates as $b_t = (x_t^{(1)}, y_t^{(1)}, \cdots, x_t^{(i)}, y_t^{(i)})^T$. According to the actual situation, a set of small regions points template patches $T(X_t^{(i)})$ are denoted as a group of 11×11 pixels windows surrounded each of tracked points to describe the point's main characteristics. $T(X_t^{(i)})$ coordinates is realized via multiplication in the homogeneous coordinates with a matrix $X_t^{(i)}$. And then let the multiple points template patches $T(X_t^{(i)})$ align to an input image $I(X_t^{(i)})$ in a time dependent micro image sequence. The most important step in tracking process is to calculate points templates with their rotational and translational features $X_t^{(i)} = (R_t, M_t^{(i)}; 0, 1)$ in each frame, where the details of R_t and $M_t^{(i)}$ will be analyzed in Section 3.1.

The efficiency of the particle filter tracking algorithm mainly relies on the importance random sampling and calculation of the weights $w_t^{(i,j)}$, where $j = 1, \cdots, G$, G is the number of sample particles. The measurement likelihood $p(y_t^{(i)}|X_t^{(i)})$ is independent of the current state particles $X_{0:t}^{(i,j)}$, the details will be shown in Section 3.2.

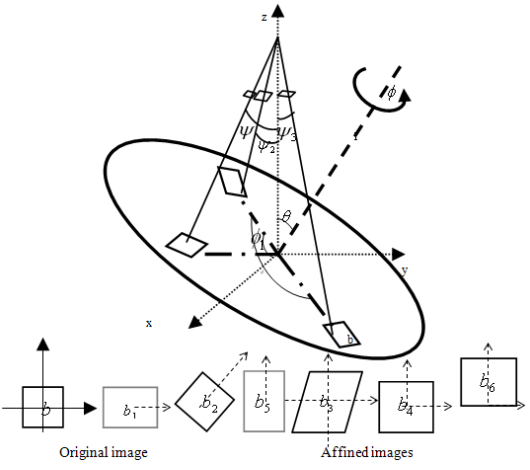


Fig. 3. The measuring micro object rotates with speed Φ on the stage of tilting angle Θ. Therefore the relative variations of camera position (denoted by Ψ) would result in the changes of projective image shape(denoted as $b1 \rightarrow b6$), which indicate the affine transformation at discrete interval time. Suppose b is an original microscopic image captured when the stage is horizontal and still, $b1 \rightarrow b6$ can be formed by b via translation, scaling and shearing transformation. The basis elements of affine transformation are represented as $\xi 1 \rightarrow \xi 3$.

3.1 Affine Motion Representation of Tracking Points

Generally, a rigid motion can be divided into a translation $M_t^{(i)}$ and a rotation R_t of the reference frame [13]. All the points of a rigid object are rotating with a same angular velocity, so the rotation $R_t^{(i)}$ of the all point templates $T(X_t^{(i)})$ can be denoted as R_t in each frame during the tracking process. The translation vectors are denoted as $M_t = (Mx_t^{(1)}, My_t^{(1)}, \cdots, Mx_t^{(i)}, My_t^{(i)})^T, i = 1, \cdots, N$, so the action of a point rigid motion can be written as

$$l(X_t^{(i)}) = R_t X_t^i + M_t^{(i)}, \tag{1}$$

where l is a nonlinear function to represent the action of points rigid motion, R_t is an invertible 2×2 rotation matrix and $R_t \in \Re^2, M_t^{(i)}$ is a $2 \times i$ translation vector and $M_t^{(i)} \in R^2$.

Each rigid motion can be represented with R_t and $M_t^{(i)}$ as a Lie group of two-dimensional affine transformations matrix [11,13], Lie group to Lie algebra [13,14] can be characterized using the exponential map, whose relationship is rewritten as the form

$$\begin{bmatrix} R & M_t \\ 0 & 1 \end{bmatrix} = \exp \begin{bmatrix} \Gamma & \Upsilon \\ 0 & 0 \end{bmatrix}, \tag{2}$$

where $\Gamma \in gl(2), \Upsilon \in \Re^2$. The basis elements $\xi_m (m = 1, \cdots, 3)$ of affine transformation can be calculated by the right equation, which are represented via (3):

$$\xi_1 = \begin{bmatrix} 0\ 0\ s_x \\ 0\ 0\ s_y \\ 0\ 0\ 0 \end{bmatrix} \xi_2 = \begin{bmatrix} s_x\ 0\ 0 \\ 0\ s_y\ 0 \\ 0\ 0\ 0 \end{bmatrix} \xi_3 = \begin{bmatrix} 0\ s_x\ 0 \\ s_y\ 0\ 0 \\ 0\ 0\ 0 \end{bmatrix}, \tag{3}$$

where s_x, s_y is the value of affine transformation unit. As shown in Fig. 3, b_1, b_5 are produced via a translation transformation as s_x, s_y respectively varying in ξ_1; b_2, b_3 are produced via shear transformation with shape deformation, whose length and width are both changed as s_x, s_y varying in ξ_2 respectively; b_4, b_6 via a scaling changing as s_x, s_y respectively varying in ξ_3.

3.2 Particle Filter Tracking Algorithm with Affine Motion Parameters

The tracked distortion templates are accurate represented and matched with affine geometric transformation in each frame is the key to points tracking in long micro image sequences.

State Estimating and Sampling for Tracked Points in Each Frame. The state equation is geometrically well-defined as

$$X_t^{(i)} = X_{t-1}^{(i)} \exp(a_i \log((X_{t-2}^{(i)})^{-1} X_{t-1}^{(i)}) + \sum_{m=1}^{3} \xi_m \varepsilon \tau_t), \tag{4}$$

where a_i is the AR process parameter. Since the velocity has been varying about this monocular micro vision rotational system, AR model can't be calculated owing to the resample process, AR-based state dynamics model can be understood as an infinitesimal constant velocity model. τ_t is the zero-mean Gaussian noise. $\varepsilon = (1/12)^{1/2}$ represents obtaining 12 frames per second in the micro image sequences. The measurement state equation can be expressed in the discrete setting as

$$y(X_t^{(i)}) = l(X_t^{(i)}) + v_t, \tag{5}$$

where v_t is a measurement zero-mean Gaussian noise.

To optimize the computational procedure and avoid directly calculating the weighted sample mean $\{X_t^{(i,1)}, \cdots, X_t^{(i,G)}\}$ for every point template, we have to firstly approximate the sample mean of $R_t^{(i,j)}$ and $\Gamma_t^{(i,j)}$, where $\{R_t^{(i,1)}, \cdots, R_t^{(i,G)}\}$, $t = 1, \cdots, N, i = 1, \cdots, k, j = 1, \cdots, G$(here j is the number of sample particles, t and i are as defined before). When we resample the particles according to their weights at every time-step, all the resample particles can be expected to be quite similar to each other. The sample mean of $X_t^{(i,j)}$ can be approximated as

$$\begin{bmatrix} \overline{R}_t^{(i)} & \overline{M}_t^{(i)} \\ 0 & 1 \end{bmatrix} = \exp \begin{bmatrix} \overline{\Gamma} & \overline{\Upsilon} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} R_{t,max}^{(i)} \frac{1}{G} \sum_{j=1}^{G} \log((R_{t,max}^{(i)^{-1}} R_t^{(i,j)}) & \overline{M}_t^{(i)} \\ 0 & 0 \end{bmatrix}, \tag{6}$$

where $\overline{M}_t^{(i)}$ is the arithmetic mean of $M_t^{(i,j)}$.

(a) The fixed spatial structure of those rigid points

(b) The point region with salient gradients information

Fig. 4. The spatial structural and regional salient information of feature points

Calculate Covariance with Descriptor for the Tracked Point Templates. The values of the covariance related with the Wiener process also play an important role in multiple points tracker performance. We extend the combined state-parameter estimation method [16], the sample covariance matrices U, calculate its minimum distance between two elements of U, which are related to covariance descriptor $S(X_{t=0}^{(i)})$.

$$U = \frac{1}{G-1}\sum_{j=1}^{G}\mu_t^{(i,j)}\mu_t^{(i,j)^T} \tag{7}$$

$$\mu_t^{(i,j)} = \log(U_t^{(i,j)}) - \frac{1}{G}\sum_{j=1}^{G}\log(U_t^{(i,j)}) \tag{8}$$

where μ is column vector. The covariance descriptor is calculated at next.

For the rigid points with fixed spatial structure, we can also add structure constraints to minimize the influence of illumination, distortion, motion blur and noise interference. Define the structure constraints vector as

$$h = (x_t^{(i)}, y_t^{(i)}, b_t^{(i,d)}, b_t^{(i,\theta)}, I(X_t^{(i)}), I_x, I_y, tan^{-1}(I_x/I_y), I_{xx}, I_{yy})^T,$$

where $(x_t^{(i)}, y_t^{(i)})$ is the center coordinate, $b_t^{(i,d)} = \sqrt{(x_t^{(i)} - x_t^1)^2 + (y_t^{(i)} - y_t^1)^2}$ and $b_t^{(i,\theta)} = tan^{-1}((x_t^{(i)} - x_t^1)/(y_t^{(i)} - y_t^1))$ respectively represent the distance and the angle between origin (or polar axis) and each tracked pixel in polar coordinate system. $I(X_t^{(i)})$ denotes the intensity of image and I_x, I_y, I_{xx}, I_{yy} represent the first- and second-order image derivatives in Cartesian coordinates system respectively [15,17]. The point templates patches covariance descriptor S can be given as

$$S_{X_{t=0}^{(i)}} = \frac{1}{s-1}\sum_{p=1}^{s}(h_{(X_t^{(i)})_p} - \hat{h})(h_{(X_t^{(i)})_p} - \hat{h})^T, \tag{9}$$

where $s = 11 \times 11$ is the area of template window, $\hat{h}$ is the mean value of $h_{(X_t^{(i)})_p}$, the similarity between $S_{X_{t=0}^{(i)}}$ and $S_{X_t^{(i)}}$ can be understood as the Log-Euclidean metric distance between themselves.

Measurements of Relative Distance. For the covariance descriptors of the tracked points templates are changing successively during the tracking process, it is necessary to collect image covariance descriptors and calculate their principal eigenvector and the geodesic distance between two group elements the $\{S(X_t^{(i)}), \overline{S}\}$ and $\{T(X_t^{(i)})), \overline{I}\}$. The measurement function can be defined using the distance-from-feature-space, distance-in-feature space (refer to [16,17]) and similarity comparison purposes, then the measurement equation as in reference [15] can be more explicitly expressed as:

$$y_t^{(i)} = \begin{bmatrix} \sqrt{\left\|\log(S_{X_t^{(i)}}) - \log(\overline{S})\right\|^2 - \sum_{n=1}^{M} c_n^2} \\ \sqrt{\sum_{n=1}^{M} \frac{c_n^2}{\rho_n}} \\ \left\|I_{X_t^{(i)}} - \overline{I}\right\| \end{bmatrix} + v_t^{(i)}, \tag{10}$$

where c_n is the projection coefficient, ρ_n is the eigenvalue, both of these two parameters are used to calculate the distance-in-feature-space, and $\overline{I}$ represents the point mean intensity. Then the measurement likelihood can be described as

$$p(y_t^{(i)}|X_t^{(i)}) \propto \exp((y_t^{(i)})^T R^{-1} y_t^{(i)}), \tag{11}$$

where R is the covariance of zero-mean Gaussian noise v_t. When the measurement likelihood $p(y_t^{(i)}|X_t^{(i)})$ is gained, we can calculate and normalize the importance weights for the tracked points and realize multiple points tracking in long microscopic sequences.

4 Experiments

With the proposed methods we discussed in Section 3, there will provide experimental results both single point and multiple points tracking on several long micro image sequences. The micro-polyhedrons image sequences (*0*, *5* and *six-ray star*) are used for testing the tracking performance of the proposed method. These sequences are consisted respectively of 627, 249, 74 frames with the discrete time interval 12, 15, 6 frames per second (fps), whose illumination direction, tilting angle and rotating rate are all changed.

To acquire long micro image sequences, micro videos are divided into successive frames. Due to the weak texture, noise disturbed and motion blur, it is hard to detect the single point locations exactly by normal methods. In order to solve these problems, at the beginning of tracking, the initial templates T(X0 (i)) were given, then let them align to $I(X_0^{(i)})$, owing to calculate R_0 and $M_0^{(i)} = (M_{x_0}^{(i)}, M_{y_0}^{(i)})$. The initial coordinate vector b_0 of tracked key points is often extracted and given before tracking starting. Throughout the testing, search windows are constant-size (11×11 pixels).

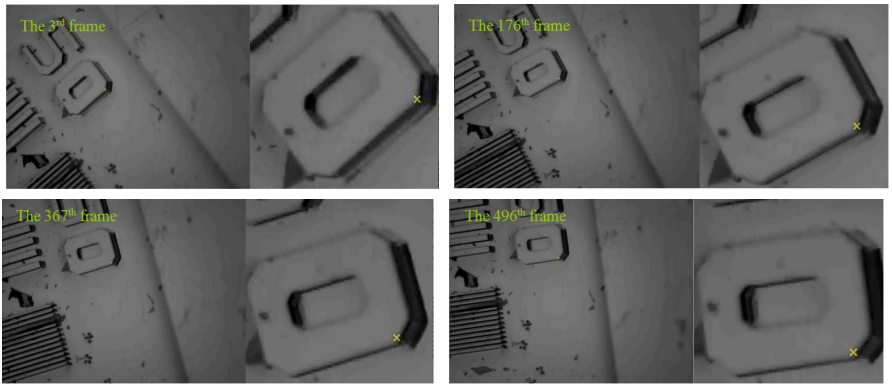


Fig. 5. Tracking results of single feature point. In each sub-figure, the left column is the original tracking results of single point, while the right column is the amplified view of the tracking result.

In Section 4.3, we also test the same micro-polyhedrons image sequences respectively with KLT tracking algorithm, then analyze and discuss the error from two perspectives, such as 3d-reconstruction and calculating the distance between the experimental results and the true coordinates.

4.1 Single Feature Point Tracking

The first experiment measures tracker high-accuracy and reliability for tracking single point in long micro image sequences, as shown in Fig.5. The maximum number of tracked points is set $k = 1$. The center coordinates $b_0 = (274, 222)^T$ and an initial rotating angle 52° are used to form the initial rotation matrix R_0 for the given template $T(X_0^{(1)})$. Given $M_0^{(i)} = b_0$, the template is updated in each 16 frames of the sequences to avoid tracking failure and a_i is empirically chosen to be 0.6. When the sampling particles are $G = 600$, the best points tracking results are gotten.

4.2 Multiple Feature Points Tracking

The second experiment measures tracker reliability and robust for tracking multiple points in the three above-mentioned long micro image sequences, some representative results are shown in Fig. 6.

To describe the details clearly, we take micro-polyhedron *0* for example. Firstly, we initialize the rotational matrix R_0 and M_0 for the rigid multiple points template patches $T(X_0^{(i)})$ and define the origin coordinate vector as $0^T = (0_t^{(1)}, 0_t^{(1)}, \cdots, 0_t^{(i)}, 0_t^{(i)})^T$ with the same number as the tracked points. M_t is translated from origin coordinate vector of images to the current point vector b_t in each frame. The translation vector $M_0 = (M_{x_0}^{(1)}, M_{y_0}^{(1)}, \cdots, M_{x_0}^{(i)}, M_{y_0}^{(i)})^T, (i =$

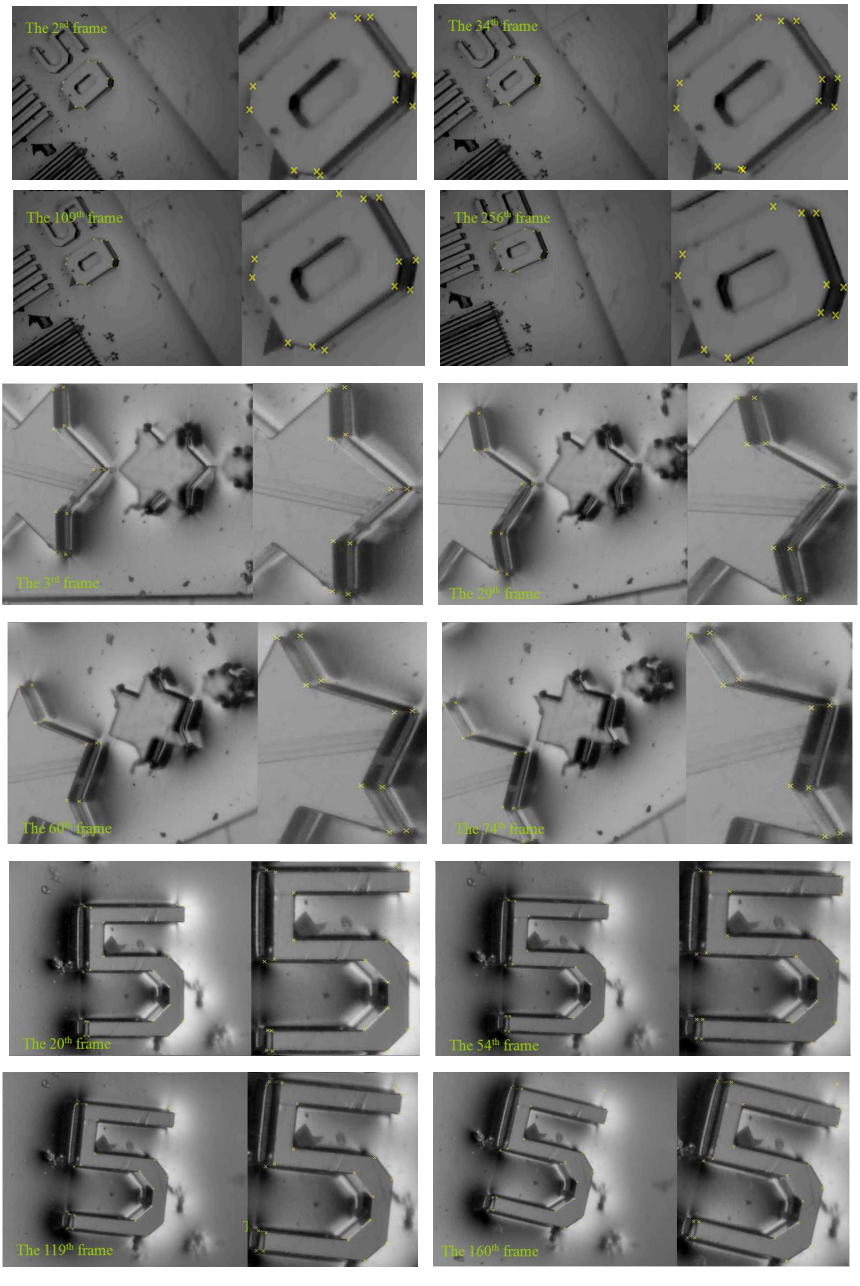


Fig. 6. PF with affine transformation tracking results of multiple key points on different micro structures: rows 1-2 are of *0* (tilt 10°, rotate $4\pi/225$ rad/s, amplified by 10×1.9 times, 640×480), rows 3-4 are of *six-ray star* (tilted 15°, rotate $\pi/90$ rad/s, amplified by 10×9.3 times, 1280×960) and rows 5-6 are of *5* (tilt 15°, rotate $\pi/45$ rad/s, amplified by 10×9.3 times, 640×480). In each sub-figure, the left column is the original tracking results, while the right column is the amplified view of the tracking results.

$1, \cdots, 12)$ is equal to $b_0 = (274, 222, 271, 200, 237, 153, 214, 150, \cdots)^T$. The templates are updated in each 16 frames of the sequences to avoid tracking failure and a_i is empirically chosen to be 0.6. After several experiments, the best points tracking results were got when the sampling particles are $G = 600$. It deserves to be specially noted that the parameters a_i are changed from 0.6 to 0.8 and the initial rotating angle 52° to 90°, the results of micro image sequences of 5 and a six-ray star are more accurate.

The results of multiple key points tracking by proposed method for microscopic frames are obviously robust and accurate. For further comparison with the state-of-the-arts tracking algorithm, we provide some tracking results by the popular KLT algorithm in the following part.

4.3 Comparative Experiments and Error Analysis

Except for directly evaluating the results in previous experiments, this section attempts to find other evaluation methods, such as comparing with the classical KLT tracking algorithm, 3D reconstruction and calculating the numerical values of error.

Tracking Experiments by KLT. We test the KLT algorithm on our micro image sequences, which are the same sequences used in the last experiments with the proposed method. We can easily tell that although KLT tracking algorithm usually performs good on tracking tasks in normal-scale rigid objects image sequences, it really has a poor performance for micro-image sequences, as shown in Fig. 7.

In the initial frames, 80,160 and 120 points are extracted respectively from micro-polyhedrons *0*, the *six-ray star* and *5*, which are mainly determined by those points contained all the key points. However, there are only 23, 26, 23 points left in the last frame with KLT algorithm tracking. From those images in Fig.7, we can clearly see the test results about several micro-polyhedrons, such as 0, 5 and a six-ray star with KLT algorithm tracking algorithm which couldn't precisely track key points in micro-image sequences. Since there are lots of points disappeared and drifted which has been shown in Fig.7. For example, in the yellow square regions in the second and third rows, it is not hard to find that there are lots of points (containing 2 key points) disappeared during the KLT algorithm tracking process. In the elliptical regions in the third row there are some points drifted, namely wrong tracking. Those factors lead to erroneous results of 3D- reconstruction.

3D Reconstruction Experiments with Tracking Results. Micro 3D- reconstruction plays an important role in computer vision, which is one of the objectives of tracking and correspondence, and can also be used to judge the tracking algorithm to be good or bad. To evaluate our work, we also reconstruct

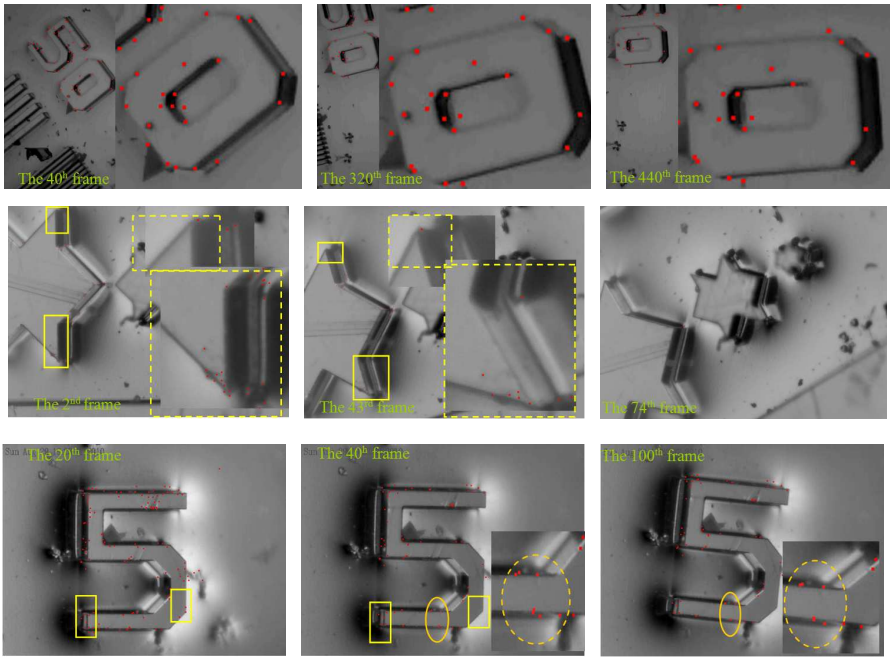


Fig. 7. Tracking results of multiple key points by KLT, where dotted lines are amplified by the same shape. row 1 are results of *0* (tilt 10, rotated 4/225 rad/s, amplified by 10x1.9 times, 640x480), the second row is of *six-ray star* (tilted 15, rotate /90 rad/s, amplified by 10x9.3 times, 1280x960), the last row is of *5* (tilt 15, rotate /45 rad/s, amplified by 10x9.3 times, 640x480).

the 3D micro-structure of micro image sequences from a single camera view. Firstly, we track respectively by PF with affine transformation algorithm and KLT tracking algorithm to acquire the tracked multiple points results of micro-polyhedrons, such as 0, the six - ray star and 5. Among the 627, 74, 249 micro image frames, the former acquires 12, 10, 21 tracking correspondences point results, while the latter acquires 23, 26, 23 tracking correspondences points results in long micro sequences. Secondly, those results were used to reconstruct the micro rigid structures 3D-reconstruction, since those sequences are taken by no-calibrated cameras, so the depth is set to 500, which is an essential parameter for 3D-reconstruction. Finally, the camera relative trajectorys to rotating micro structures were either sketch in Fig.8.

In fig.8, we can draw the conclusion that KLT algorithm tracks multiple points of micro image sequences are much worse than PF with affine transformation tracking algorithm. Since the former cannot follow the location of micro key points precisely and cannot gain the complete micro-structure, while the latter can track key points accurately and robustly.

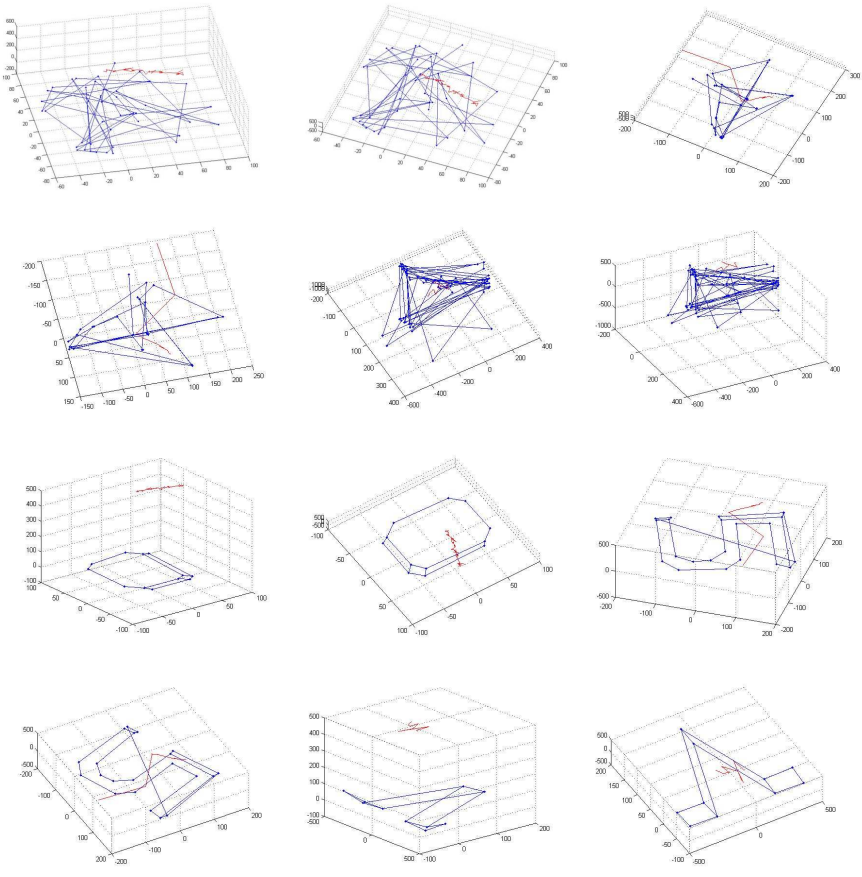


Fig. 8. The comparison of 3D-reconstruction results using the tracking points by PF and KLT respectively. The top two rows using the datas of KLT algorithm tracking multiple points results, while the following two rows using the datas of PF with affine transformation tracking algorithm multiple points tracking results. The irregular red curves in the figure represent the camera relative trajectory.

Tracking Errors Analysis. We calculate the mean square errors between the tracked points coordinates $(x_t^{(i)}, y_t^{(i)})$ and the true coordinates (x, y) as

$$err = \frac{1}{k}\sqrt{\sum_{i=1}^{k}((x_t^{(i)} - x)^2 + (y_t^{(i)} - y)^2)}, \tag{12}$$

where $(x_t^{(i)}, y_t^{(i)}), i = 1, k, t = 1, \cdots, 627$ are acquired by calculate templates geometric center and a set of (x, y) are the true coordinates which can be obtained by manual tracking. In the two figures, there are several pixels errors which mainly caused by those micro-images themselves which have blurred and strong

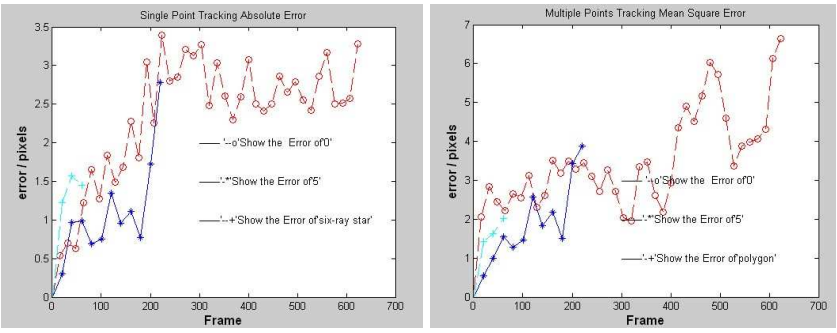


Fig. 9. The errors of points tracking results of PF combined with affine transformation algorithm

edges as well as the similar color and textures in surrounding regions. Those factors on one hand increase the tracking difficulty, on the other hand result in some artificial errors when choose the true coordinates.

The tracking errors are analyzed by calculating the distances between those tracked points coordinates and their true coordinates. In Fig.9, the left is single point tracking absolute errors. Its maximum error is lower than 3.3 pixels, which shows the tracker is robust and accurate to point track compared with the results of [11]. In Fig.9, the right is multiple points tracking mean square errors, which show the mean square errors are saliently higher than single point tracking error when $t > 400$ frame, this mainly results from the point self-occlusion during the rotating process. The weak feature points which are almost the same color as the background or without clear gradient information would be tracked lost in the long micro image sequences show a secondary cause.

5 Conclusions

In this paper we presented a novel multiple points tracking method for a long microscopic sequences. We test the proposed method on our micro stereo imaging system. The tilting rotation of the stage produces an affine geometric transformation on the projection of fixed rigid spatial microstructure. We applied an adapted and accurate multiple feature points tracking algorithm based on particle filter (PF) with affine transformation. The experimental results show the feasibility of proposed method for rigid multiple key points tracking in long microscopic sequences, the comparative experiments shows that our method outperforms classical KLT algorithm on tracking in microscopic sequences. In the future we would like to investigate the method of decreasing tracking error in multiple points tracking case. The self-occlusion effects of feature points during the rotating process will be the major problem to be dealt with in the further study.

References

1. Brandenburg, B., Zhuang, X.: Virus Trafficking-learning from Single-virus Tracking. Nat. Rev. Microbiol. 5(3), 197–208 (2007)
2. Yilmaz, A., Shafique, K., Shah, M.: Target Tracking in Airborne Forward Looking Infrared Imagery. Image and Vision Computing 7(21), 623–635 (2003)
3. Barth, A., Franke, U.: Where Will the Oncoming Vehicle be the Next Second? In: IEEE Intell.Veh. Symp., pp. 1068–1073 (2008)
4. Martínez, E., Torras, C.: Qualitative vision for the guidance of legged robots in unstructured environments. Pattern Recognition 8(34), 1585–1599 (2001)
5. Li, B., Meng, Q., Holstein, H.: Reconstruction of segmentally articulated structure in freeform movement with low density feature points. Image and Vision Computing 10(22), 749–759 (2004)
6. Chen, X., Zhou, X., Wong, S.T.C.: Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy. IEEE Trans. on Biomedical Engineering 4(53), 762–766 (2006)
7. Ngo, T.D., Le, D.-D., Satoh, S., Duong, D.A.: Robust Face Track Finding in Video Using Tracked Points. In: IEEE International Conference on Signal Image Technology and Internet Based Systems, pp. 59–64. IEEE Press, Bali (2008)
8. Luo, Z., et al.: Feature Tracking Algorithms Based on Two Cameras. Journal of Computer-Aided Design and Computer Graphics 7(14), 646–650 (2002)
9. Ye, L., Wang, Y.: Grid Real-time Tracking of the Shoot Point from Light Pen Based on Camshift. In: IEEE International Conference on Intelligent Networks and Intelligent Systems, pp. 560–564. IEEE Press, Wuhan (2008)
10. Yao, Y.-S., Chellappa, R.: Dynamic Feature Point tracking In an Image Sequence (EKF). In: ICPR 1994, pp. 654–657 (1994)
11. Buchanan, A.M., Fitzgibbon, A.W.: Combining local and global motion models for feature point tracking. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1–8. IEEE Press, Minneapolis (2007)
12. Kwon, J., Lee, K.M., Park, F.C.: Visual tracking via geometric particle filtering on the affine group with optimal importance functions. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 991–998. IEEE Press, Miami (2007)
13. Soatto, S., et al.: Motion estimation dynamic vision. IEEE Transactions on Automatic Control 41(3), 393–414 (1996)
14. Kwon, J., Park, F.: Visual tracking via particle filtering on the affine group. In: Proc. IEEE, ICIA 2008, pp. 997–1002 (2008)
15. Kwon, J., Park, F.C.: Visual Tracking via Particle Filtering on the Affine Group. The International Journal of Robotics Research 29, 198–217 (2010)
16. Baker, S., Matthews, I.: Lucas-Kanade 20 Years On:A Unifying Framework. International Journal of Computer Vision (56), 221–255 (2004)
17. JPorikli, F., et al.: Covariance tracking using model update based on lie algebra. In: IEEE Conf. on Computer Vision and Pattern Recognition, pp. 728–735. IEEE Press, New York (2006)

Meshless Simulation of Plastic Deformation and Crack

Liang Zeng, Bo Wu, Jiangfan Ning, Jiawen Ma, and Sikun Li

School of Computer, National University of Defense Technology,
410073 Changsha, China
wubogfkd@gmail.com

Abstract. We present a meshless method for local deformation and crack simulation of plastic thin shell. Although previous meshless methods have done the similar simulations, there exists a problem that the moment matrix will be singular when the simulation points are co-planar or co-linear. The consequence is that the shape function cannot be constructed to finish the simulation and special work is needed to deal with the problem. In this paper, we propose a meshless method – Local Radial Basis Point Interpolation Method (LRPIM) to carry out the simulation. The shape function is constructed using the radial basis function (RBF) and polynomial basis function, and it guarantees that the moment matrix is nonsingular without any other assistance. Results show that our method is feasible and effective.

Keywords: meshless method, deformation, crack, RBF, LRPIM.

1 Introduction

Many physical phenomena, such as fire, smoke, water, solid deformation and fracture [1-5], are needed to be simulated to enhance the realistic immersion. Especially, the realistic simulation of material deformation and crack [2, 6] plays an important role in strengthening interactivity and reality in many fields, such as computer games, animated films, surgery simulation and virtual reality. The demands for the effects in movies and games have encouraged many researches on the physically based modeling of deformation and fracture in recent years.

In the field of deformation and crack simulation, there are three main methods: mass-spring, finite element and meshless. Mass-spring method [7] is simple and easy to be implemented; however, tuning a mass-spring system to get a certain behavior is a difficult task. Moreover, it is also difficult to accurately simulate the motion of material using mass-spring system. Finite element method (FEM) [6, 8] could obtain the accurate results. However, complex remeshing is required which makes the method unstable. Meshless method [9, 10] has been used for the simulation in computer graphics for just a couple of years. Using this method, we need only to consider little about the connection of vertices of the objects.

Previous researches [6, 11, 12] on fracture simulation concern brittle fracture, which treats objects as rigid material. However, plastic material is widely used in the real world, thus it is important to simulate the deformation of plastic objects. O' Brien

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 127–137, 2012.
© Springer-Verlag Berlin Heidelberg 2012

et al. [8] proposed the ductile fracture simulation using FEM for the first time. But there are few papers talking about the ductile fracture simulation of plastics material using meshless methods. In this paper, we concern the local deformation and crack simulation of plastic material. Friendly interactivity is also an important factor that should be considered. Our major contributions can be summarized as follows:

1. We novelly introduce a meshless method -- LPRIM into deformation and crack simulation of plastic material in computer graphics. LPRIM is used to approximate the displacements at all points and it can fix the problem of singular moment matrix in other meshless methods.
2. We propose a simple but effective dynamics resampling method, which supports multi-local deformation and crack simulation of plastic objects.

2 Related Work

Terzopoulos et al. [1] innovatively achieved physical animation of deformable objects by solving the underlying elasticity equations. Soon after, the work was extended for plastic materials and a technique for modeling viscoelastic and plastic deformations was presented [2]. O' Brien et al. [6] first introduce the FEM into graphical animation to simulate the fracture of brittle objects. They handled the fracture by using element cutting and dynamic remeshing technique. The model is extended to simulate plastic deformations and ductile fracture effects with strain state variables [8]. Virtual node algorithm was presented by N. Molino et al. [11] to automatically cut the tetrahedral mesh and form an embedded triangulated surface. The above methods are all mesh-based method, which are time-consuming and unstable.

Meshless method was first introduced into computer graphics by Desbrun and Cani [9], in which they used particle system to model soft inelastic materials with a smooth iso-surface. M. Desbrun et al. [10] used the smoothed particles to represent sample points, which enable the approximation of the values and derivatives of local physical quantities inside a medium. In the reference [13], SPH is used to model deformable solids. M. Muller et al. [14] presented a method for modeling and animating objects from stiff elastic material to highly plastic material. They use a moving least squares procedure to compute the spatial derivatives of the discrete displacement field. M. Pauly et al. [15] proposed a new meshless animation framework for solids fracture based on M. Muller. Recently, N. Liu et al. [16] proposed a meshless framework based on MLPG to simulate the fracture of brittle material.

In this paper, we propose the LRPIM meshless method to simulate the deformation and crack of plastic thin shell. In the simulation, the displacements, strains and stresses of the simulation points are computed at first. Then use the shape function to approximate the displacements of computation points according to the corresponding simulation points. The cracks at the points are determined by the maximal eigenvalues of the stress tensor. The results reveal our method is feasible and effective.

3 Meshless Simulation

In the meshless methods, the shape function is the core part for approximating the solution of PDEs. In our meshless method, the radial basis function (RBF) and polynomial basis function are used to construct the shape function for the simulation. Combining a low degree polynomial basis function, it guarantees that the moment matrix is non-singular. At the same time, our method is a truly meshless method as compared to some of the so-called meshless or element-free finite element methods. As for simulation of plastic material, we first introduce some theories about continuum mechanics as follow.

3.1 Continuum Mechanics

As a plastic object lies in three dimensions space, shown in Fig. 1, Ω indicates the field of the problem. Γ is its boundary surface and it has two parts: Γ_t is the natural boundary and Γ_u is the displacement boundary. And **b** is the vector of body force and **n** is the outward normal of the boundary [21]. When the external force acts on the object, the force will unevenly distribute on the boundary and in the interiority. The force will result in the deformation and there will be a displacement field in the solid, which will introduce stress to countervail the external force. The deformation and crack simulations are determined by the displacement field. Depending on the structure and the boundary condition, the displacements at distinct points are usually different.

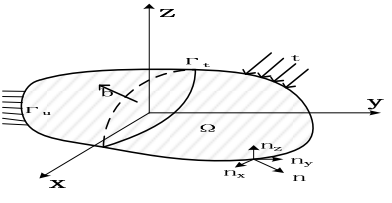


Fig. 1. A 3D continuum solid

Usually, the stress state of any point in the solid could be described via six independent stress components.

$$\boldsymbol{\sigma}^{\mathrm{T}} = \begin{bmatrix} \sigma_{xx} & \sigma_{yy} & \sigma_{zz} & \sigma_{xy} & \sigma_{yz} & \sigma_{xz} \end{bmatrix} \tag{1}$$

Where σ_{ij} indicates the stress acts on the i surface, and the direction is j. Correspondingly, there are six independent strain components on any point. The strain is the deformation ratio of the object. The equation (2) shows the relationship between strain and displacement.

$$\boldsymbol{\varepsilon}=\mathbf{Lu} \tag{2}$$

In the equation **ε** is the strain at the point, **u** is the displacement vector and **L** is the differential operator. Assume the material as Hookean material, then the following

equation describes the relationship between the stress **σ** and the strain **ε,** D is the matrix of material attributions.

$$\boldsymbol{\sigma}=\mathbf{D}\boldsymbol{\varepsilon} \tag{3}$$

According to the theories, the equilibrium equation (4) reveals the relation between strain ε and stress σ.

$$\mathbf{L}^{\mathrm{T}}\boldsymbol{\sigma}+\mathbf{b}=\rho\ddot{\mathbf{u}}+c\dot{\mathbf{u}} \tag{4}$$

In the equation, ú is the vector of velocity and ü is the vector of acceleration. If we only consider the case of statics and combine the equation (2) and (3), the equation (4) can be rewritten as follow:

$$\mathbf{L}^{T}\mathbf{D}\mathbf{L}\mathbf{u}+\mathbf{b}=\mathbf{0} \tag{5}$$

The equation (6) is the boundary condition.

$$\boldsymbol{\sigma}\mathbf{n}=\mathbf{t} \quad \text{on} \quad \Gamma_t \quad \text{and} \quad \mathbf{u}=\bar{\mathbf{u}} \quad \text{on} \quad \Gamma_u \tag{6}$$

Where **t** is the vector of the surface tractions, **ū** is the vector of the displacements.

3.2 LRPIM-Based Displacement Approximation

During the deformation and crack simulation, in order to determine where may be broken, we should compute the displacements at all the points. Then calculate the strain at the points responding to the displacements. According to the Hooke law, we will compute the stress at every point. Then calculate the eigenvalues of the stress tensor. If the maximal eigenvalue of stress tensor exceeds the material threshold, it will trigger the crack. The following flow chart shows the process. To accurately calculate displacements at all the points are unpractical and will cost too much space and time. Besides, for the computer graphics field, it is unnecessary to calculate so exactly like continuum mechanics.

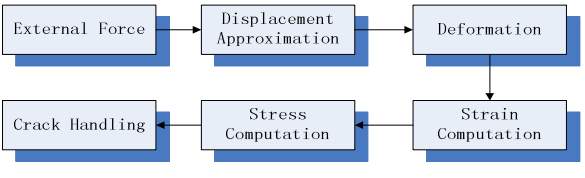


Fig. 2. The flow of computation

Assume there are n points whose displacements are given; an advisable method is to derive the other points' displacements from the given n points. In other words, how to use the known n points to approximate the other points is our main technique issue. In meshless methods, the function to achieve the approximation of the given function is called shape function. In the following section, we will introduce how to construct the shape function of LRPIM. The Fig. 3 shows the basic idea of approximating the given function in meshless method.

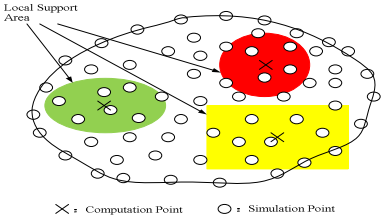


Fig. 3. Different local support areas of computation points

In Fig. 3, the forks indicate the computation points whose displacements are needed to be calculated. The circle in red, ellipse in green, rectangle in yellow are the local support areas of the computation points; the small circles indicate the simulation points, whose displacements are already known. Then the displacement of computation point is approximated via the corresponding values of simulation points which lie inside the local support area of the computation point.

Shape Function Construction. As described above, shape function used to approximate the given function value is usually with the form as equation (7) in meshless method. $\Phi_i(x)$ is the shape function at the i simulation point and $\mathbf{u}_i$ is the given function value at point i.

$$u(\mathbf{x}) = \sum_{i=1}^{n} \phi_i(\mathbf{x})\mathbf{u}_i \tag{7}$$

Point interpolation method is one of the simplest approaches to construct the shape function of meshless methods and often used to approximate the integral functions or differential functions. There are two main different point interpolation methods: polynomial point interpolation method and radial basis point interpolation method. The polynomial point interpolation method is one of the most important interpolation method and widely used in many areas, such as FEM. A series of polynomial basis functions are used to approximate the value of any points in the domain.

$$u(\mathbf{x}) = \sum_{i=1}^{m} p_i(\mathbf{x})\mathrm{a}_i = \mathbf{p}^T(\mathbf{x})\mathbf{a} \tag{8}$$

Where $\boldsymbol{p}_i(\mathbf{x})$ is the polynomial basis function in spatial coordinates, a_i is the corresponding coefficient. The polynomial point interpolation method owns some good properties and is easily implemented. However, if $\mathbf{p}^T$ is a singular matrix, there may be problem and we cannot ensure the values of coefficients. To avoid the problem, we adopt the radial basis functions to construct the shape function of the meshless method.

Similar to polynomial point interpolation method, the equation of radial basis point interpolation method is shown in equation (9).

$$u(\mathbf{x}) = \sum_{i=1}^{n} R_i(\mathbf{x})a_i + \sum_{j=1}^{m} p_i(\mathbf{x})b_j = \mathbf{R}^T(\mathbf{x})\mathbf{a} + \mathbf{p}^T(\mathbf{x})\mathbf{b} \tag{9}$$

Where $R_i(\mathbf{x})$ is the radial basis function, n is the amount of simulation points; $\mathbf{p}^T(\mathbf{x})$ is the series of polynomial basis functions, m is number of polynomial basis functions. a_i and b_j are the corresponding coefficients. The basis function is not necessary, but it will bring some advantages at little cost of computation. It will make the approximation more accurate and the system more stable.

As for the radial basis function $R_i(\mathbf{x})$, it is a real-valued function whose value depends only on the distance from some other points. The sums of radial basis function are usually used to approximate given functions. Commonly used types of radial basis functions include Gaussian, MultiQuadric, Inverse MultiQuadric and so on. We adopt the following Gaussian function in our program.

$$R_i(r_i) = R_i(\| \mathbf{x} - \mathbf{x}_i \|) = e^{[-a_c(\frac{\|\mathbf{x}-\mathbf{x}_i\|}{d_c})^2]} \tag{10}$$

Where $\| \bullet \|$ indicates the Euclidean distance, a_c is the shape parameter, d_c is the eigen-distance in the local support area of point $\mathbf{x}$. Since the $R_i(\mathbf{x})$ is always greater than zero, it will avoid problems with ill conditioning of the matrix solved to determine coefficients. According to [17], Gaussian radial basis and linear polynomial basis function together guarantee the moment matrix is non-singular. So it avoids the singularity of moment matrix in other meshless methods.

As the displacements of the n simulation points are known, we will derive equation (11) from equation (9).

$$U_n = R_n a + P_m b \tag{11}$$

Where U_n is the vector of displacements at the n simulation points. R_n is the symmetric n×n matrix of the radial basis function and P_m is the vector of polynomial basis function at the n simulation points. Since the displacements and coordinates at the n simulation points are already known, there are (n+m) unknowns but the number of equation is only n. In order to make the equations have a unique solution, we add the new constraints as follow.

$$\sum_{i=1}^{n} p_j(x_i) a_i = P_m{}^T a = 0 \quad , \quad j=1, 2, \cdots, m \tag{12}$$

Combine equation (11) and equation (12), we will get equations with (n+m) unknowns and (n+m) equations. Then we will obtain the values of the coefficients as follow:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} R_n & P_m \\ P_m^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} U_n \\ 0 \end{bmatrix} \tag{13}$$

The above inverse matrix composed of Gaussian function and the polynomial basis function is guaranteed to be non-singular. As we get the values of the coefficients, the equation (9) could be rewritten as follow.

$$u(x) = \begin{bmatrix} R^T(x) & p^T(x) \end{bmatrix} \begin{bmatrix} R_n & p_m \\ p_m{}^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} U_n \\ 0 \end{bmatrix} \tag{14}$$

Contrast with equation (7), we will find that:

$$\sum_{i=1}^{n}\phi_i(\mathrm{x}) = \begin{bmatrix} R^{\mathrm{T}}(\mathrm{x}) & p^{\mathrm{T}}(\mathrm{x}) \end{bmatrix} \begin{bmatrix} R_n & p_m \\ p_m{}^T & 0 \end{bmatrix}^{-1} \tag{15}$$

So $\sum\Phi_i(\mathrm{x})$ is the shape function of LRPIM. Then the displacements at all the points in the Ω field can be calculated.

Displacement of Simulation Point. In the previous sections, we assume that the displacements at the n simulation points were already known. Actually, for any object, the displacement at any point is usually unknown and hardly to measure. The following is the process to compute the displacements of simulation points. The local weak form of equation (5) is as follow:

$$\int_{\Omega_q} W_I(\sigma_{ij,j} + b_i)d\Omega = 0 \tag{16}$$

Where W_I is the weight function for the local support area of point I and usually is the spline function. According to Liu [18], the equation can be rewritten as follow:

$$(\mathrm{K}_I)_{3\times 3n}(\mathrm{u})_{3\mathrm{n}\times 1} = (\mathrm{f}_I)_{3\times 1} \tag{17}$$

Since all the parameters are already known, the integrations can be calculated. We use Gaussian integration to compute their values in our implementation.

4 Deformation and Crack Algorithm

Dynamic Resampling. The point-based geometry is used in our simulation. One of its advantages is that the resampling operation is easily implemented. For point-based geometry, the point is the basic element for surface representation and the density of the points is a key factor for the quality of rendering. If the number of points is not large enough, some details of the surface will be lost, especially in the local area which we care about. Adversely, we will get more details with less artificial. So the adequate density of points should be acquired. Assume the point density of the initial model are sufficient, however, the density will change during the simulation. Then the dynamics resampling operation is needed to make sure the surfaces of objects are well sampled. The dynamics resampling makes the computation of displacement approximation more accurate and makes the surface representation better.

During the initialization, we maintain a list of neighbouring particles for every point. The point and its neighbouring particles form a watertight surface. During the simulation, the displacements at all the points may change and it will affect the distance between them. If the distance between two neighbour points is greater than the diameter of the surfel, there may be flaw on the surface. To keep the surface watertight, new surfels should be added. The Fig. 4 shows the new surfel insertion process. The attributions of the new surfel, such as location, normal and color, are interpolated by its parent surfels. After inserting the new surfels to the surface, we should add the points to the neighbouring list and update information for corresponding points.

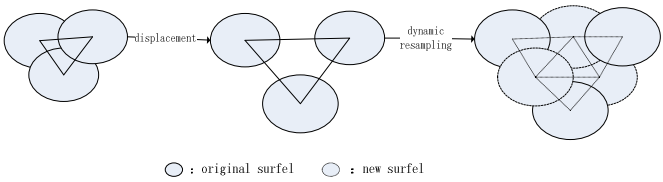


Fig. 4. New surfels are inserted during the simulation

Deformation and Crack Simulation. We propose a simple and effective method to simulate the deformation and crack for the objects. The method is based on dynamics resampling. As for the deformation simulation, once the simulation points transfer from original location to others, the computation points whose support areas include the simulation points will also be displaced. The displacements of the computation points are calculated via the LRPIM. If the distance between two neighbouring points is greater than the diameter of the surfel and neither satisfies the crack condition, the dynamics resampling operation will be executed and update the information of neighbouring list at the same time.

For the crack simulation, if the simulation points give birth to displacements, then the displacement of the computation points whose support areas include the simulation points must be calculated by the LRPIM. And once the maximal eigenvalue of stress tensor is greater than the material threshold, the crack will happen at the point. Then if the distance from it to any point, which is in the crack point's neighbouring list, is greater than the diameter of the surfel, then remove the point from the neighbouring list. Therefore, although the two points are adjacent in geometry, the dynamics resampling won't be executed even if they apart from each other and there will be crack on the surface.

5 Rendering

In our experiment, we actually use splatting to represent the surface of object instead of triangle mesh. The splatting technique was proposed by Zwicker [19]. We use splatting technology to render the point sets as continuous surfaces. It is a simple and efficient approach for high-quality representation of point-sampled surfaces, which uses a z-buffer algorithm to resolve visibility. The points can be processed efficiently without any additional acceleration data structures by splatting. It supplies a way to project the sampled points to pixels. The differences between direct points rendering and splatting are shown in Fig. 5.

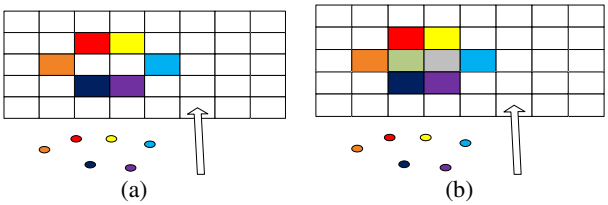


Fig. 5. Direct rendering and splatting

The direct points rendering, as in Fig. 5a, just assign the attributes of the points to the closest pixel. Therefore, there may be some holes in the image for the pixels without corresponding points. In Fig. 5b, using splatting will get a continuous surface, as the "holes" in Fig. 5a will be assigned the attributes according to its neighboring pixels. The attributes of the "holes" are calculated using the following formulation.

$$f(p) = \sum_i w_i r_i (p - p_i) \tag{18}$$

where p indicates the pixel without corresponding projected point and p_i indicates the pixel on which some points are projected; r_i is the basis function of the distance between p and p_i; w_i is the attribute value of p_i. Then the attributes of p, such as color and normal, can be evaluated from the given point sets.

6 Implementation and Results

In our implementation, the geometry models are needed to do some pre-treatments to make sure the point sets are sufficient for the rendering and calculating. At the same time, we should build a neighbouring list to reserve the information about adjacent points. For radial basis function, we adopt the Gaussian function which makes the matrix invertible in most situations. The geometry models are rendered on Visual Studio2008 using OpenGL. The computation of matrix is based on the Newmat library. The following figures show our results.

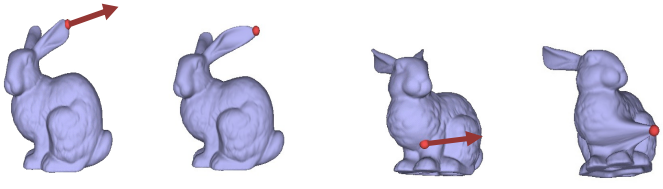


Fig. 6. Deformation at different local areas

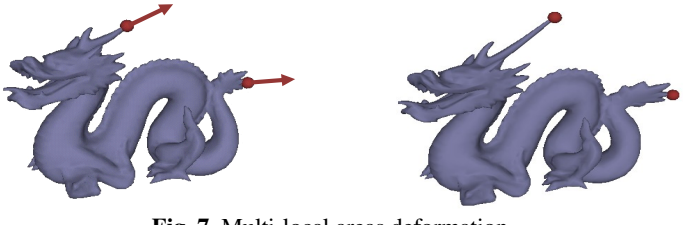


Fig. 7. Multi-local areas deformation

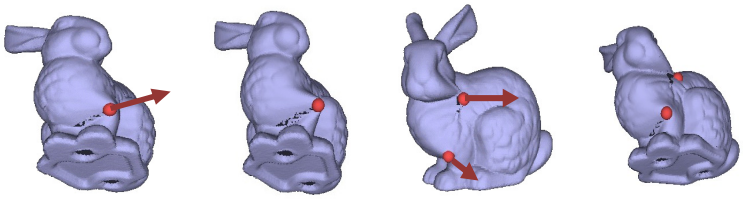


Fig. 8. Crack at local areas

The above figures show our results. For the dragon model, the number of vertices is 14.99k. 122 simulation points are chosen in a random way to make the simulation. The vertices number of the bunny model is 34.84k and the number of simulation points is 324.

In Fig. 6, the deformations of bunny are shown. The little red ball indicates the place where the external force acts on. When the traction acts on the ear of the plastic bunny, the ear is elongated. The last two images show the similar results when the traction acts on different part. The Fig. 7 shows the results that more than one external force act on the dragon model.

In Fig. 6 and Fig. 7, we only simulate the deformation of models and do not check whether the maximal eigenvalue of stress tensor exceeds the threshold of the material. The Fig. 8 shows the results of deformation and crack of bunny. If the maximal eigenvalue does not exceed the threshold, the model only deforms, or else the model will crack. The lift two in Fig. 8 shows the crack simulation when the force is single and the right two shows the multi-areas cracks. In all the simulations, even if the chosen simulation points are co-planar or co-linear, the moment matrix will still be non-singular and the shape function can also be constructed to finish the simulations.

7 Conclusion

We have introduced a meshless method--LRPIM for the plastic thin shell simulation of deformation and crack. Taking advantage of Gaussian radial basis function, we avoid the problem with ill condition of the matrix solved to achieve coefficients. According to the continuum mechanics and meshless theory, the displacements at the simulation points are calculated first. Then the displacements of the other points are approximated and the strain and stress are calculated. The displacements are used to compute the deformation, and the crack is determined according to the maximal eigenvalue. We adopt the splatting technique to represent the surface. It works well with meshless method. The point-based representation of model surface facilitates the control of the topology and makes the dynamic resampling efficient.

There are several interesting directions for future research. As we know, the calculation takes place at all the points. If we want to achieve accurate results, more simulation points are needed. In other words, there will be much more computation and it is a hard work for CPU. Taking advantage of GPU, we will transfer the computation process onto GPU in future, which is suitable for parallel computing. In addition, we plan to extend our method to other physical simulations.

Acknowledgments. The work is supported by the National Natural Science Foundation of China under Grant No.60873120 and the National Grand Fundamental Research 973 Program of China under Grant No.2009CB723803.

References

1. Terzopoulos, D., Platt, J., Barr, A., Fleischer, K.: Elastically deformable models. In: Computer Graphics Proceedings. Annual Conference series, pp. 205–214. ACM SIGGRAPH (1987)

2. Terzopoulos, D., Fleischer, K.: Modeling inelastic deformation: viscolelasticity, plasticity, fracture. In: Proceedings of the 15th Annual Conference on Computer Graphics and Interactive Techniques, pp. 269–278. ACM Press (1988)
3. Nguyen, Q.D., Fedkiw, R., Jensen, W.H.: Physically based modeling and animation of fire. In: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, pp. 721–728. ACM Press (2002)
4. Fedkiw, R., Stam, J., Jensen, W.H.: Visual simulation of smoke. In: SIGGRAPH the 29th Annual Conference on Computer Graphics and Interactive Techniques, pp. 721–728. ACM Press (2002)
5. Feldman, E.B., O'Brien, F.J., Arikan, O.: Animating suspended particle explosions. In: Proceedings of ACM SIGGRAPH, pp. 721–728 (2003)
6. O'Brien, F.J., Hodgins, K.J.: Graphical modeling and animation of brittle fracture. In: Proceedings of ACM SIGGRAPH, pp. 287–296 (1999)
7. Smith, J., Witkin, A., Baraff, D.: Fast and controllable simulation of the shattering of brittle objects. In: Graphics Interface, pp. 27–34 (2001)
8. O'Brien, F.J., Bargteil, W.A., Hodgins, K.J.: Graphical modeling and animation of ductile fracture. In: Proceedings of ACM SIGGRAPH, pp. 291–294 (2002)
9. Desbrun, M., Cani, P.M.: Animating soft substances with implicit surfaces. In: Computer Graphics Proceedings, pp. 27–34. ACM SIGGRAPH (1995)
10. Desbrun, M., Cani, P.M.: Smoothed particles: A new paradigm for animating highly deformable bodies. In: 6th Eurographics Workshop on Computer Animation and Simulation 1996, pp. 61–76 (1996)
11. Molino, N., Bao, Z., Fedkiw, R.: A virtual node algorithm for changing mesh topology during simulation. ACM Transactions on Graphics, 385–392 (2004)
12. Bao, Z., Hong, J., Teran, J., Fedkiw, R.: Fracturing rigid materials. IEEE Transactions on Visualization and Computer Graphics, 370–378 (2007)
13. Becker, M., Ihmsen, M., Teschner, M.: Corotated SPH for deformable solids. In: Proceedings of Eurographics Workshop on Workshop on Natural Phenomena (2009)
14. Muller, M., Keiser, R., Nealen, A., Pauly, M., Gross, M., Alexa, M.: Point based animation of elastic, plastic and melting objects. In: Eurographics/ACM SIGGRAPH Symposium on Computer Animation, pp. 141–151 (2004)
15. Pauly, M., Keiser, R., Adams, B., Dutre, P., Gross, M., Guibas, J.L.: Meshless animation of fracturing solids. ACM Transactions on Graphics, 957–964 (July 2005)
16. Liu, N., He, X., Li, S., Wang, G.: Meshless simulation of brittle fracture. Computer Animation and Virtual Worlds 115–124 (2011)
17. Schaback, R., Wendland, H.: Characterzation and construction of radial basis functions. In: Multivariate Approximation and Application. Cambridge University Press (2000)
18. Liu, R.G., Gu, T.Y.: An introduction to meshfree methods and their programming. Springer Press (2005)
19. Zwicker, M., Pfister, H., Baard, V.J., Gross, M.: Surface Splatting. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (2001)

Lossless Visible Three-Dimensional Watermark of Digital Elevation Model Data

Yong Luo[1,2], Yan Zhao[2], Lei Cheng[2], Jianxin Wang[1], and Xuchong Liu[1]

[1] School of Information Science and Engineering, Central South University, Changsha 410083, China
[2] College of Science, National University of Defense Technology, Changsha 410073, China
yngluo@163.com

Abstract. Digital elevation model (DEM) data describe the information of ground elevation. So it is important to protect the copyright of digital elevation model data. A lossless visible 3-D watermarking algorithm to protect DEM data is proposed in this paper. The copyright watermarking is embedded in the DEM data by an visible way. The original data, blocked by 3-D visible watermarking, are hidden in the watermarked DEM data by a generalized histogram algorithm. Because the visible watermark blocks a part of DEM data, illegal users are restricted to retrieve. At the same time, 3-D visible watermark can identify the copyright. Without original 3-D watermark data, authorized users can eliminate the 3-D visible watermark and restore the original DEM data lossless by applying the proposed algorithm. It is a blind watermarking algorithm. Experiments demonstrate that the proposed algorithm has satisfactory security and can effectively protect the copyright of DEM data.

Keywords: Visible watermarking, DEM data, Lossless.

1 Introduction

DEM (Digital Elevation Model) is a digital representation of elevation information. It is a reconstruction of sample data of true terrain. DEM theory has achieved great developments since its foundation. DEM has become an important part of the digital earth massive database and the basic framework of three-dimensional virtual reality.

As DEM data describe ground elevation information, they are widely used in mapping, hydrology, meteorology, geomorphology, geology, soil, construction, communications, meteorology, military and other areas of the national economy and national defense. Therefore, it is necessary to protect the copyright and security of DEM data.

This paper presents a three-dimensional visible digital watermarking for protection of DEM data. The 3-D digital copyright information is embedded into DEM data as a visible watermark. The visible digital watermark can indicate the copyright, and restrict unauthorized users from retrieving key information of protected DEM data. On the other hand, the visible 3-D digital watermark can be erased, and original DEM

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 138–147, 2012.
© Springer-Verlag Berlin Heidelberg 2012

data can be restored losslessly. This method is capable to satisfy high precision requirements.

Typical digital watermarking techniques use signal processing methods, and embeds visible or invisible watermarks into carriers. For invisible watermarking, users can extract the hidden watermark to certify copyright; while for visible digital watermarks, they can be removed by using the hiding information to gain full content of the protected data. Digital watermarking technique enables effective protection of digital rights and information security. As the DEM data record the true terrain elevation data, it is not allowed to be destroyed. Therefore it is necessary to study the lossless digital watermarking method.

Researchers propose a number of lossless watermarking algorithms. C. W. Honsinger[1] presents a lossless watermarking algorithm, which restores an original image containing embedded data by adding 256 module hash values into the original image. The disadvantage of this algorithm is to introduce salt and pepper noise. Fridrich proposed RS method [2],which would produce gray value overflow. Celik [3]proposes a generalized LSB algorithm. Tian [4] propose an expansion difference method, which hides secret information in the difference of pixel gray. But both of them lose some amount of information hidden. Zhicheng Ni et al proposed lossless watermarking based on modified histogram method [5]. This method can hide more information, and enable high-quality of carrier image. At the same time, it has low computational complexity. However, in these methods, the watermarking is embedded into an image by invisible way. Because the DEM data are the true elevation, these image watermarking algorithms cannot be directly applied to DEM data.

Some papers [6,7] studied the DEM watermarking. The method of paper [6] is loss watermarking, so it is not proper method since data accuracy is important for DEM. An invisible lossless DEM watermarking algorithm is proposed in the paper [7].The original DEM data is similar to the watermarked DEM data. It cannot restrict illegal users, because they are able to obtain the all data with small errors. Therefore, security for DEM data is not satisfied. In addition, some researchers study the triangle mesh data[8-15]. But they are all lossy and invisible watermarking. Equidistant meshing data(DEM) is our research object. This paper proposes a lossless visible watermarking algorithm for DEM data.

To our knowledge, a lossless DEM visible watermarking technique is not available yet. The lossless watermarking based on modified generalized histogram can protect DEM data effectively. The generalized histogram of DEM data is established, and shifts to hide the data are used to eliminate the visible 3-D watermark. Legal users can extract hidden information to eliminate the visible 3-D watermark and restore the DEM data lossless.

2 DEM Data Generalized Histogram

The digital elevation model data are constructed by sampling or interpolating uniformly-spaced, and they can be written as follow in matrix form.

$$\begin{bmatrix} Z_{00} & Z_{01} & \cdots & Z_{0(n-1)} \\ Z_{10} & Z_{11} & \cdots & Z_{1(n-1)} \\ \cdots & \cdots & \cdots & \cdots \\ Z_{(n-1)0} & Z_{(n-1)1} & \cdots & Z_{(n-1)(n-1)} \end{bmatrix} \tag{1}$$

Z_{ij} is the terrain altitude in grid node i, j. It is the actual terrain features with the spatial distribution model. As we know, the range of terrain altitude on the earth is in [-14000, 9000].

Firstly, we fused the visible 3-D watermark with the DEM data to obtain the watermarked digital elevation data. Set watermarked DEM data as I, the size is $N_1 \times N_2$, and the access value marked $Z_{ij} \in [m, M]$. M and m are the float-point data. A part of DEM data are blocked by the visible watermark. Set blocked DEM data as W.

$$W = \{w_i \mid i = 0,1,2,...,P\} \tag{2}$$

They are the hiding information then we divide $[m, M]$ into n small intervals, and the length of interval is h.

$$h = \frac{M - m}{n} \tag{3}$$

These intervals are marked as $\{I_k = [m + hk, m + h(k+1)], k = 0,1,2,...,n-1\}$ k is the interval index, which is defined as generalized gray. The number of altitude data locate in $[m + kh, m + (k+1)h]$ is signed $h(k), k = 0,1,2,...,n-1$. Through scanning all the DEM data, we can get the generalized gray value and histogram $H(I) = \{h(k) \mid k = 0,1,2,...,n-1\}$. DEM data and the generalized histogram are shown in figure 1.

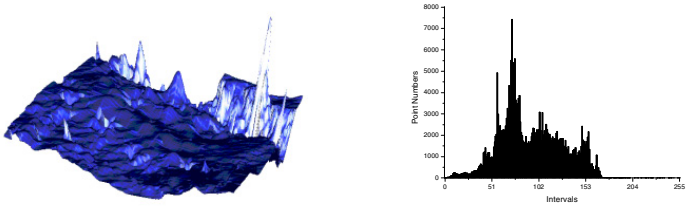


Fig. 1. DEM data and the generalized histogram

3 Lossless 3-D Visible Watermarking

The section describes the algorithm of information hiding and recovery processing. Figure 2 is the flow chart of information hiding, and figure 3 is the flow chart of information lossless recovery.

Now, We describe the information hiding process. Firstly, protected data and 3-D watermark are fused to get the watermarked digital elevation model data I. Thus a

part of digital elevation model data is blocked by 3-D watermark, these information are recorded as W. W is the hiding information. Secondly, we scan the watermarked digital elevation model data I to obtain generalized histogram $\{H(k) \mid k = 0,1,2,...,n-1\}$. Fnally, we adopt generalized histogram shifting algorithm to hide the blocked information W into I.

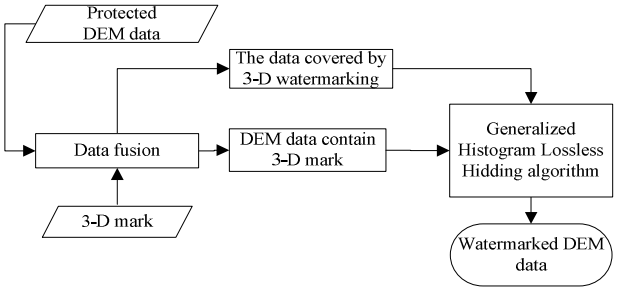


Fig. 2. Information hiding procedure

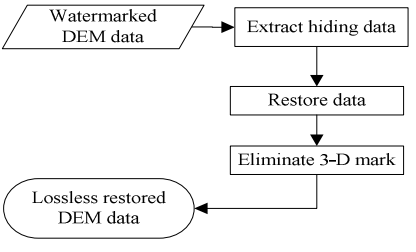


Fig. 3. Lossless restoring procedure

Recovery procedure is shown in figure 3. Firstly, we extract hidden information, and restore the blocked digital elevation model data W . Finally, W takes place of the visible 3-D watermark in same position . So the DEM data is restored lossless.

4 Lossless Information Hiding Algorithms

Set the value of each coordinate point $Z_{ij} \in [m, M]$, where m and M are the minimum and maximum of DEM data. Scan the watermarked DEM data I to produce generalized histogram $H(I)$. The interval $[m, M]$ divided into n small intervals. From left to right are recorded as $\{I_k, k = 0,1,2,...,n-1\}$. he length between each district is h. Set the maximum value of generalized histogram $H(I)$ is $h(a)$, $a \in [0, n-1]$. Without loss of generality, we can assume that

$$\sum_{k=0}^{a_1-1} h(k) > \sum_{k=a_1+1}^{n-1} h(k) \quad . \tag{4}$$

That is, the number of coordinates of point contained in intervals $\{I_k \mid k = 0,1,...,a-1\}$ is fewer than the intervals $\{I_k \mid k = a+1, a+2,...,n-1\}$. Watermark embedding steps are as follows:

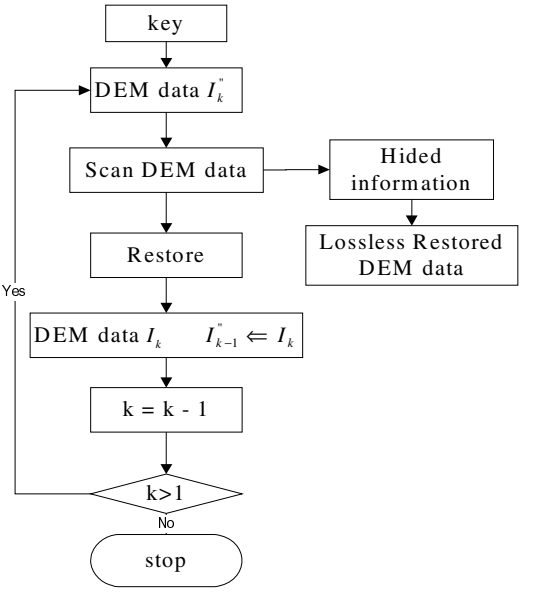
Step1: All the intervals in $H(I)$ of which index $k \in [a+1, n-1]$ move right one interval. In other words, add h to the altitude of pixels whose index is in internal $k \in [a+1, n-1]$. Then we get a new digital elevation model data I' which has no pixels in the a+1 interval. Mark the moving direction with Sand $s = 0$, represent moving left, $s = 1$ represent moving right. In this case, we move right, so $s = 1$.

Step2: Scan data I', and search the pixels whose altitudes Z_{ij} drop in intervals I_a. That is

$$m + (a-1)h \le Z_{ij} < m + ah \tag{5}$$

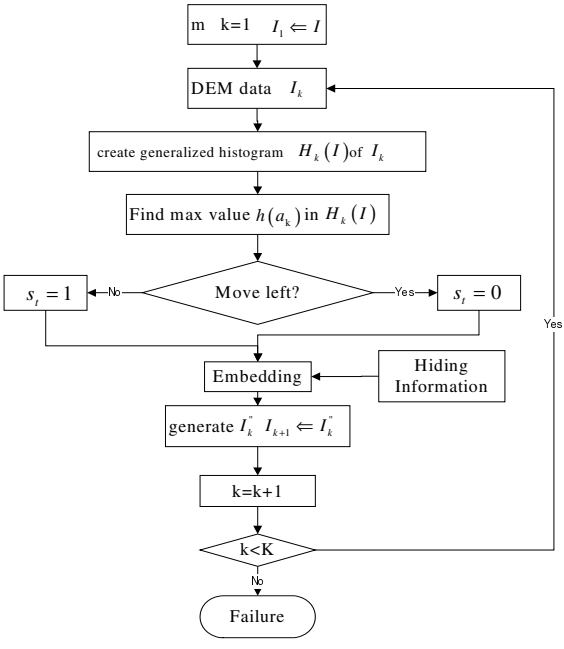
Then compare the watermark information. If the watermark w_k is a bit of information 1, add h to the altitude Z_{ij}; if the watermark w_k is a bit of information 0, keep the pixel it alone.

$$\begin{cases} Z'_{ij} = Z_{ij} + h, when\ w_k = 1 \\ Z'_{ij} = Z_{ij}, when\ w_k = 0 \end{cases} \tag{6}$$



(a)

Fig. 4. The embedding flow chart



(b)

Fig. 4. (*continued*)

After the above two steps, we get the DEM data $I^{"}$ which contain 3-D watermark. The lowest altitude m in the original DEM data I, the interval length h, the index of embedding interval a, and moving direction s are the key to extract watermark from $I^{"}$.

Extracting watermark is the reverse process of embedding it. The watermark extraction steps are as follows:

Step1: Divide the DEM data $I^{"}$ which contain the watermark. For $s = 1$, so split altitude to n+1 intervals from the minimum m by step h. Then scan $I^{"}$ by pixels, if encounter altitude in the a+1 interval, extract watermark bit "1"; if encounter altitude in an interval, extract watermark bit 0

$$\begin{cases} m+(a-1)h \le Z_{ij}^{'} < m+ah, w_k = 0 \\ m+ah \le Z_{ij}^{'} < m+(a+1)h, w_k = 1 \end{cases} \tag{7}$$

Step2: Restore the DEM data according equation(8), where $m+(a-1)h \le Z_{ij}^{'} < m+(a+1)h.$

$$\begin{cases} Z_{ij} = Z_{ij}^{'}, when\ m+(a-1)h \le Z_{ij}^{'} < m+ah \\ Z_{ij} = Z_{ij}^{'} - h, when\ m+ah \le Z_{ij}^{'} < m+(a+1)h \end{cases} \tag{8}$$

Step3: Scan $I^{''}$ again and subtracts h from the altitude of the pixels locate in interval $[a+2,n]$.

$$Z_{ij} = Z_{ij}^{'} - h, when\ Z_{ij}^{'} \ge m+(a+1)h \tag{9}$$

Now, original DEM data is recovered without loss. The real data embedded capacity is $P = h(a)$. If the length of embedded information is greater than h(a), we choose other maximums

$$h(a_1), h(a_2), \cdots, h(a_k) \tag{10}$$

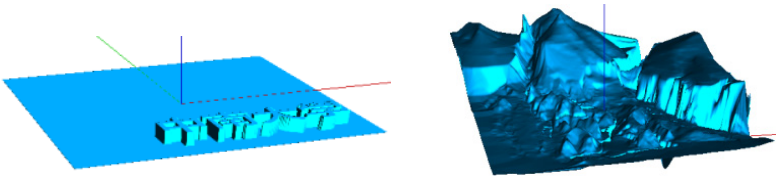
They are used to hide left blocked DEM information. We get DEM data $I_k^{''}$ which contained all the watermark The extracting key include the lowest altitude m, the interval length h, embedding interval indexes $a_1, a_2, \cdots a_k$, the moving directions $s_t \left(\in \{0,1\}, t = 1,2,\cdots,k\right)$ and the remained capacity $[h(a_1)+h(a_2)+\cdots+h(a_k)]-P$. As a matter of fact, the embedded capacity reaches up to 10000 bits which is enough for the DEM data. The equation (11) are the information which are recorded to restore the DEM data.

$$\begin{cases} a_1, a_2, ..., a_k \\ s_1, s_2, ..., s_k (s_i = 0\ or\ s_i = 1, i = 1,2,...,k) \\ [h(a_1)+h(a_2)+\cdots+h(a_k)]-P \end{cases} \tag{11}$$

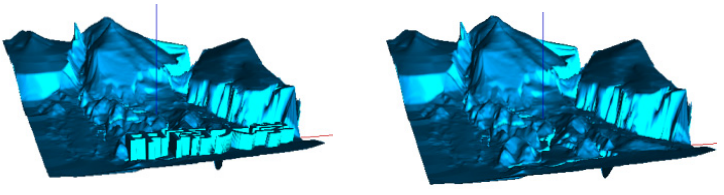
5 Experimental Result

In experiment, carriers are several digital elevation model data , which are the actual elevation for the lattice 512×512. The elevation Z_{ij} is single precision floating point. The interval $[m, M]$ is divided into 256 intervals.

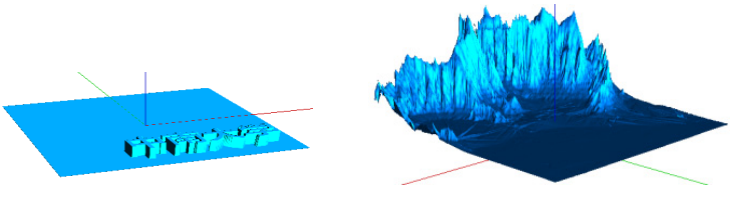
Figure 5 is the digital elevation model data for the three-dimensional visible lossless watermark experiment. Figure 5.a.1 and 5.b.1 are the three-dimensional watermark, but also a digital elevation model data. Figure 5.a.2 and figure 5.b.2 are the original digital elevation model data. The figure 5.a.3 and figure 5.b.3 are visible watermarked digital elevation model data, which contain the embedded information.



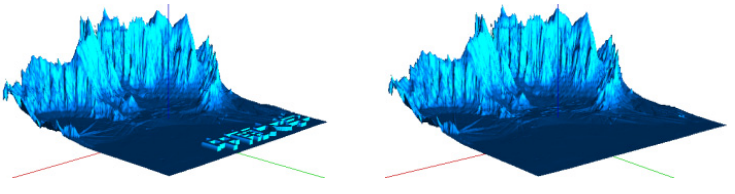
(a.1) 3D watermarking (a.2) Original DEM data



(a.3) DEM data embedded visible watermarking (a.4)Lossless restored DEM data



(b.1)3D watermarking (b.2) Original DEM data



(b.3) DEM data embedded visible watermarking (b.4)Lossless restored DEM data

Fig. 5. Experiment of lossless visible watermarking

Through extracting hidden information which are used to eliminate the visible 3-D mark, we can restore the original digital elevation model data. The lossless restored digital elevation model data are shown in Figure 5.a.4 and Figure 5.b.4. From this experiment, the watermarked digital elevation model data restored lossless. For a part of digital elevation model data blocked by the visible watermark, the illegal users are restricted. So the copyright of digital elevation model data are protected.

The visible 3-D DEM watermarking proposed in this paper first time. The watermarking algorithms proposed in paper[1-5,8-15] cannot dealing with the DEM data, and the DEM watermarking in paper[6,7] are invisible ones. Therefore, the algorithm cannot compare their performance directionally. On the other hand, this watermarking algorithm has the disadvantage that it is not able to against the attacks. It is a fragile watermarking algorithm. When the watermarked DEM data is under

attack, the order of the extracting information is destroyed. So the block digital original data cannot restore correct. Of course, the purpose of this algorithm is to limit the use of illegal users. From this perspective, the algorithm has its applications.

6 Conclusion

This paper presents a lossless three-dimensional visible watermarking, which can be used to protect the DEM data. For digital elevation being floating-point data, the generalized histogram is designed to embed hidden information. This algorithm has low computational complexity. It is easy hardware implementation. Experiments demonstrate that three-dimensional watermark is able to be eliminated, and the digital elevation data are restored lossless. We can draw the conclusion that this visible 3-D lossless watermark can protect DEM data effetely.

This visible 3-D watermark can restrict the illegal users. But it is wake to against attacks. In order to further protect the DEM data, the next step is to improve the ability of the watermark against attacks.

References

1. Li, L., Pan, Z.G., Zhang, M.M., Ye, K.: Watermarking subdivision surfaces based on addition property of Fourier transform. In: GRAPHITE 2004, pp. 46–49 (2004)
2. Li, L., Zhang, D., Pan, Z.G., Shi, J.Y., Zhou, K., Ye, K.: Watermarking 3D mesh by spherical parameterization. Computers & Graphics 28(6), 981–989 (2004)
3. Yin, K.K., Pan, Z.G., Shi, J.Y., Zhang, D.: Robust mesh watermarking based on multiresolution processing. Computers & Graphics 25(3), 409–420 (2001)
4. Tian, J.: Reversible Watermarking by Divergence Expansion. In: Proceedings of Workshop on Multimedia and Security: Authentication, Secrecy, and Steganalysis, pp. 19–22 (2002)
5. Ni, Z.C., Shi, Y.Q., Nirwan, A., Wei, S.: Reversible Data Hiding. IEEE Transactions on Circuits and Systems for Video Technology 16(3), 354–362 (2006)
6. Mi, H., Cheng, L.Z., Luo, Y.: Lossless watermarking for digital elevation mode data. Computer Engineering and Applications 43(30), 40–43 (2007)
7. Luo, Y., Cheng, L.Z., Cheng, B., Wu, Y.: Study on digital elevation mode data watermark via integer wavelets. Journal of Software 16(6), 1096–1103 (2005)
8. Cayre, E., Macq, B.: Data hiding on 3D triangle meshes. IEEE Transactions on Signal Processing 51(4), 939–949 (2003)
9. Ming, L., Bors, A.G.: Shape watermarking based on minimizing the quadric error metric. In: Proceedings of IEEE International Conference on Shape Modeling and Applications, Beijing, China, pp. 103–110 (2009)
10. Hu, R., Rondao, P., Macq, B.: Constrained optimization of 3d polygonal mesh watermarking by quadratic programming. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, China, pp. 1501–1504 (2009)
11. Wang, W.B., Zheng, G.Q., Yong, J.H., Gu, H.J.: A numerically stable fragile watermarking scheme for authenticating 3D models. Computer-Aided Design 40(5), 634–645 (2008)

12. Benedens, O.: Geometry-based watermarking of 3D models. IEEE Computer Graphics and Applications 19(1), 46–55 (1999)
13. Ohbuchi, R., Masuda, H., Aono, M.: Data embedding algorithms for geometrical and non-geometrical targets in three-dimensional polygonal models. Computer Communications 21(15), 1344–1354 (1998)
14. Cho, J.W., Prost, R., Jung, H.Y.: An oblivious watermarking for 3-D polygonal meshes using distribution of vertex norms. IEEE Transactions on Signal Processing 55(1), 142–155 (2007)
15. Wang, W.B., Zheng, G.Q., Yong, J.H., Gu, H.J.: A numerically stable fragile watermarking scheme for authenticating 3D models. Computer-Aided Design 40(5), 634–645 (2008)

3D Articulated Hand Tracking Based on Behavioral Model

Zhiquan Feng[1], Bo Yang[1], Yi Li[2], Haokui Tang[1], Yanwei Zheng[1], Minming Zhang[3], and Zhigeng Pan[4]

[1] School of Information Science and Engineering, University of Jinan, Jinan, 250022, P.R. China
[2] School of Control Science and Engineering, University of Jinan, Jinan, 250022, P.R. China
[3] State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou, 310058, P.R. China
[4] Hangzhou normal university, Hangzhou, 310002
Shandong Provincial Key Laboratory of Network Based Intelligent Computing, Jinan, 250022
{ise_fengzq,Yangbo,Liyi,ise_tanghk,ise_zhengyw}@ujn.edu.cn, zmm@cad.zju.edu.cn, zhigengpan@gmail.com

Abstract. Taken it into consideration that human has a great deal of experiences and knowledge of hand postures, if these operating skills of postures are applied to HCI, the simple and convenient human-computer interface can be expected. In fact, tracking, recognition and interaction based on 3D freehand are a part of the cores in our virtual assembly system, but it is a challenging task to track 3D freehand in real-time because of high dimensionality of 3D full hand model. A novel framework for 3D freehand tracking is put forward in this paper. Firstly, we model and investigate this problem under our virtual assembly system (VAS), so as to decrease the arbitrariness and complexity of this issue. Secondly, we put emphasis on building cognitive and behavioral model (CBM) for users in VAS. Thirdly, we research on the way to track 3D freehand based on CBM. The main contributions of this paper are that we propose a new CBM, TPTM model, provide a way to connect users and computer for effective interaction, and present a real-time freehand tracking algorithm. Based on TPTM model, the prediction, the number of particles, the way and scope of sampling, are optimized. TPTM model not only explain behavioral characteristics for users but also can effectively guide the design of freehand tracking algorithm. TPTM model also provides a data structure that can facilitate the implementation of the tracking algorithm. Our experimental results show that the proposed approach raises the quality of each sampled particle or avoid sampling "poor" particles which appear with low probability in each frame, and it tracks 3D freehand in real-time with high accuracy. The number of the drawn particles is reduced up to 5 and the tracking speed increase up to 81 ms per frame.

Keywords: 3D Freehand Tracking, cognitive behavioral models, Human-computer Interaction.

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 148–165, 2012.
© Springer-Verlag Berlin Heidelberg 2012

1 Introduction

There has been a great emphasis lately in HCI (Human Computer Interaction) research to create easier and more natural to use interfaces by making direct use of natural communication and manipulation skills of humans. The hand can technically be regarded as a device with more than 20 DOFs, it forms the most effective, general purpose, interaction tool for HCI. Skill learning systems, surgical simulations, and robot instruction or virtual environments in general are several advanced applications requiring direct sensing of hand and/or finger motion [3]. Hand posture is natural, intuitionistic and easy to learn for human-computer interaction. Using direct hands as the input, media are not needed any more for communication between human and computer, and users can control a computer by means of an appropriate defining posture. Hand posture is also a non-verbal intercommunion way between human and human, for example pointing direction, moving an object and expressing human emotion. In fact, recent trends in user's interfaces have brought on a wave of new consumer devices that can capture the motion of our hands. These include for behaviour understanding [21], sign language recognition [8], pointing mouse [6], controlling virtual objects [24], controlling TV [18], controlling windows commands [19], drawing pictures [6], controlling robot [7]. In Virtual Reality (VR) system, wearing data gloves to interact with computer impacts on immersion as well as convenience, and so tracking, recognition and interaction based on 3D freehand are important problems. Despite a very high level of interest in the computer vision community, the general freehand tracking problem remains largely unsolved and currently markerless trackers still cannot compete in accuracy, speed and robustness with commercial motion capture systems.

Motivated with the need of human-computer interface in our virtual assembly system, we attempt to track 3D freehand in real time with satisfactory accuracy. The difficulties of this issue lie in the following aspects. Firstly, the hand is a flexible object, its projection results in a large variety of shapes with many self-occlusions. Secondly, it remains challenging to recover 3D structure from a hand image in computer vision domain. Thirdly, freehand is a typical articulated object with high dimensionality, and to find out the real 3D hand model from nearly unlimited hand gestures is almost impossible. Furthermore, the issue of obtaining observation values automatically and robustly from frame images sequence is still a problem to be further researched.

2 Related Work

2.1 Hand Tracking

Visually capturing human hand motion requires estimating the 3D hand global pose and its local finger articulations, which is a challenging task due to high dimensionality of full hand kinematics, fast movements and frequent self-occlusions.

The particle filter (PF) has attracted more and more attention in articulated object tracking as an approach to addressing non-Gaussian and non-linear problems. The drawback of PF tracker is that the high dimensionality of the hand structure causes the exponential increase of the number of particles [12]. Another issue of PF tracker is that, prediction is performed by using the history of the hand states and requires identifying a system model. Modelling the non-linear dynamics of the hand motion is not an easy problem [3].

The first method to deal with high dimensionality is to decompose state space into sub-spaces [HT03, DDR01], and then PF is imposed respectively on each sub-spaces. The second method is to obtain the relationship between hand postures and high-dimensional feature vector sets by means of machine learning [33]. Tony Lin et al. [29] presented the RNC-based manifold learning method for nonlinear dimension reduction. J. Deutscher et al. [9] used a continuation principle based on annealing to introduce the influence of narrow peaks in the fitness function. Based on gradient descent with adaptive and parameter specific step sizes. M. Bray [4] put forward the so-called SMD method (Stochastic Meta-Descent) to solve high-dimensional problem. Lin J. et al. [20] proposed the divide and conquer approach to estimate both global and local hand motion.

M. Bray et al. [5] extended the SMD algorithm with new features for fast and accurate tracking by adapting the different step sizes within video frames and by introducing a robust cost function which incorporated both depths and surface orientations. Raskin et al. [27] combined the annealed particle filter body part tracker with the Gaussian process dynamical model to reduce the dimension of the state vector. Y. Wu et al. [31] presented the model-based method to capture the hand articulation by learning hand natural constraints. M. Kato et al. [17] proposed an efficient representation of hand motions by ICA, and ICA basis vectors represented local features, each of which corresponded to the motion of a particular finger. Xinyu Xu et al. [32] proposed the tracking algorithm based on Rao-Blackwellised particle filter (RBPF), estimated some of the state variables as in a regular particle filter, and the distributions of the remaining variables are updated analytically using Kalman filtering. By N. Vaswani [30], at any time, most state change occurred in a small number of dimension, while the change in the rest of the state space was small.

Ankur Agarwal and Bill Triggs [1] proposed a novel approach to modeling the non-linear and time-varying dynamics of human motion, using statistical methods to capture the characteristic motion patterns that exist in typical human activities. Qing Chen et al. [25] described a new approach to solve the problem of real-time 3D hand tracking and motion analysis with a combination approach of statistical and syntactic analysis. Daubney et al. [11] used a bottom-up approach to detect instances of the object in each frame, these detections are then linked together using a high-level priori motion model. Their method entirely depended on motion. Xinxiao Wu et al. [33] proposed a novel dimension reduction method to learn the low-dimension intrinsic motion manifold of articulated objects.

Recently, Robert Y. Wang and Jovan Popović [26] put forward a glove-based method with the idea that a distinctive glove simplified the pose inference to the extent that the computer can largely determine the pose of a hand from a single frame.

We achieve hand tracking with real time by researching on C*ognitive* B*ehavioral Model* (CBM) for users on our virtual assembly platform.

2.2 CBM

Cognitive psychology is concerned with information processing, and includes a variety of processes such as attention, perception, learning, and memory. It is also concerned with the structures and representations involved in cognition, while cognitive models have been successfully used in three main ways by human-computer interaction (HCI) [16]. On the other hand, mental model has been studied by cognitive scientists as part of efforts to understand how humans know, perceive, make decisions, and construct behavior in a variety of environments. The relatively new field of HCI has adopted and adapted these concepts to further the study in its main area of concern. For most cognitive scientists today, a mental model is an internal scale-model representation of an external reality. It is built on-the-fly, from knowledge of prior experience, schema segments, perception, and problem-solving strategies. A mental model contains minimal information. It is unstable and subject to change. It is used to make decisions in novel circumstances. A mental model must be "runnable" and able to provide feedback on the results. Humans must be able to evaluate the results of action or the consequences of a change of state. They must be able to mentally rehearse their intended actions. Cognitive scientists often use academic studies of mental models to gain information on the processes of the mind.

3 Our Proposed Method

3.1 Overview

In order to deal with high dimensionality problem and track 3D freehand in real time, we firstly research on CBM for users as the breakthrough point, then design the tracking algorithm based on the CBM model. We design our algorithm under PF tracker framework, put focus on how to predict hand postures, how to draw samples and how many particles of hand postures to be drawn for PF tracker, and achieve the goals of 3D freehand tracking with real-time and high accuracy. Accordingly, the framework of the proposed method is shown in Figure 1.

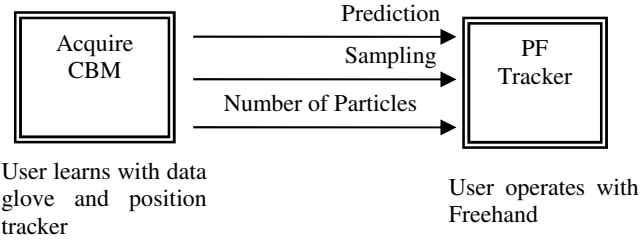


Fig. 1. The overview of the proposed method. PF tracker is based on CBM.

In the process of acquiring CBM, data glove and position tracker are needed to wear on user's hand, while in the process of freehand tracking, any hardware device or marks on user's hand are not requested any more.

3.2 VAS

The VAS is defined as a virtual assembly system with hand-driven human-computer interface, and is described as VAS={Objects, Tasks, Operations, Constraints}, among which 'Objects' stands for the objects and their relationships in the assembly scene, 'Operations' stands for the basic operations imposed on the movable objects. 'Tasks' refers to the purpose or what the user should perform, and 'Constraints' means assembly constraints including geometry constraints, scale constraints and sequence constraints. One of our basic ideas is that VAS is regarded as a system composed of cognitive tasks. In this paper, we just only consider one cognitive task------to move a 3D sphere from one place to another. All experiments, including acquisition of CBM and implementation of PF tracker, are performed in VAS.

Table 1. The description of cognitive tasks used in VAP

No.	Name	Description
0	Reset	Hand gesture backs to the initial state
1	Move	Hand move without changes of finger angles
2	Grab	The fingers bend
3	Insert	Hand puts one object into another
4	Release	The fingers extend

We will investigate cognitive and behavioral models for users in VAS.

3.3 Cognitive and Behavior Model for Users

Mental Model

For HCI practitioners, a mental model [22] is a set of beliefs about how a system works, an explanation of users' thought process about how something works in the real world, it is the most fundamental psychology process including sense, perception, attention, knowledge description, memory, learn, language, problem solution, ratiocination, imagination and so on. It attempts to explore how humans know, perceive, make decisions and construct behavior. Mental models result from people's tendency to explanations of things in the world. The field of HCI seeks to understand the explanations and hypotheses that people form about the systems that they use.

We use task analysis method [13] to capture user's mental model. For specific assembly tasks, the user tends to perform them with the lowest cognitive burden, the lowest energy cost and a most pleasure style. In order to reach these goals, the user has to from a stable cognitive and behavioral model. The core of the CBM is the way to connect the user and computer. Our study showed that there are many factors to affect interaction between the user and computer, but cognitive information is the key in our VAS system (See Fig. 2).

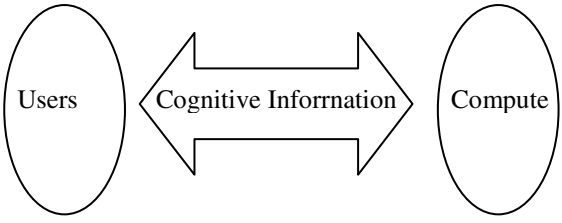


Fig. 2. The mentality model for users in VAS. The media for interaction between users and computer is cognitive information.

Cognitive and Behavior Model (CBM)

Mental model to great extent decides the behavioral style and model according to cognitive psychology. So, based on the mental model, we shape CBM with the following points.

(1) Task. VAS is regarded as a system composed of cognitive tasks (Table 1). A cognitive task can be performed by cognitive efforts of users. The procedure from the current state to destination state in the mental model is mapped to the task in our CBM model. So, the given state, destination state and constraints are basic components of a task. A task always corresponds to a dynamic hand gesture.

(2) Phase. In order to reduce cognitive burdens on user, the user always implements an entire cognitive task by means of dividing a task into several phases, which is one of the most distinctive features to perform the cognitive task. One of the advantages of phasing is that it reduces the complexity of dealing with a cognitive task because each phase is easier to be understood and performed. In many cases, phasing is a natural experience or skill coming from learning, training or exercising.

For example, in order to move an object in VAS, the user firstly translates his/her hand towards the object, then gets ready for the hand posture to grasp the object, the third phase is to grasp the object, and the last phase is to move the object to the destination position.

With the purpose of reducing dimensionality of hand motion, a phase is featured with following points:

(a) For any time of all variables of the 3D hand model, global rotation, global transformation and local motion are NOT allowed to simultaneously appear.

(b) A phase corresponds to certain semantics.

Here, global motion refers to the motion of the palm while the fingers keep unchangeable relative to the palm; local motion refers to the motion of the fingers while the palm keeps unchangeable relative to world coordinate system. For example, in a phase with semantics ‘to move’, bending or extending of fingers should not appear.

(3) Trajectory. Trajectory is the core of CBM, it explores spatio-temporal motions of hands when the user performs cognitive tasks. Each variable in hand posture vector corresponds different trajectories for each frame. The spatio-temporal model of each variable can be learned with the help of data glove and position tracker, which is also called model curve in this paper.

The trajectory is featured with the following points:

(a) For the same cognitive task, the trajectory is regarded as a random variable for each frame, it follows multiple modal distribution: uniform distribution and Gaussian distribution.
(b) The larger the jitter of a point on a trajectory, the larger the indeterminacy and unpredictability.
(c) A trajectory is described in form of

$$C = (No_T, No_P, V, \alpha_1, \alpha_2, ..., \alpha_K) \tag{1}$$

In which No_T, No_P and V are task No., phase No, and viewpoint respectively. $(\alpha_1, \alpha_2, ..., \alpha_K)$ is parameters of the curve. K refers to the length of frames.

Figure 3 is the probability distribution of the random variables from our 20 experiments for the same task. We noticed that these variables do not follow Gaussian distribution. Another interesting observation is that the point with the probability distribution is not the average point on the probability distribution curve. For example, in Fig. 3, the average value is 36 degrees, but the value with the probability distribution is 27 degrees. Further more, data analysis showed that the probabilities at turn points are very small, so, the larger the jitter of a point on a trajectory, the more difficult it is predicted.

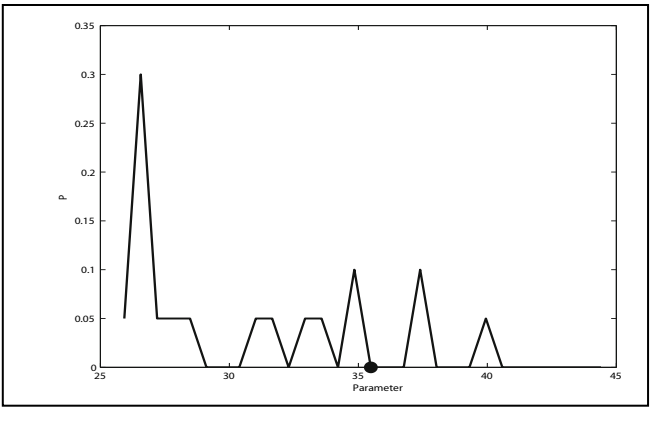


Fig. 3. The probability distribution of a variable of the 20 experiments of the same variable in hand posture model. The data comes from the data glove and position tracker. The solid circles on horizontal axis refer to the average positions of the variables.

(4) Mentality. Memories, motivation, psychological feelings, cognitive evaluation, recall and so on are the interior driving forces for a specific cognitive task. Cognitive information is a bridge linked all of these mental factors. Mental model also reflects the optimized strategy of psychological requirements and cognitive burden. The task, the phase and the trajectory can be explained with the proposed mentality model.

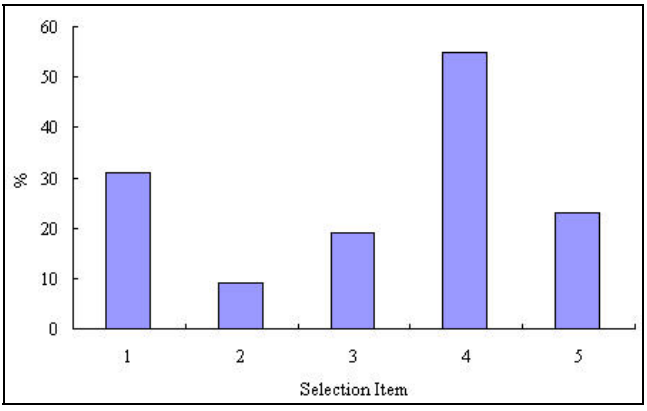


Fig. 4. The responses to a questionnaire of the psychological feelings. 1-not comfortable. 2-happy. 3-tired. 4-not natural. 5-ridiculous or ludicrous. We investigated 100.

In order to investigate the extent to which mental model plays a role in affecting the user's behavior, we investigated 100 students for issue of nature for hand behavior with questionnaire. The problem was stated as follows: when moving your hand close to an object to be grasped, your hand will not do it by a style of cruising or vibrating, what is your psychological feelings for this case? The answer is shown in Figure 4.

The point of this experiment is that the nature of hand behavior is the main driven force for the operator to perform a task. The lowest cognitive burden, the lowest energy cost and a most pleasure style are some of the fundamental demands in shaping CBM. So, tracking natural hand based on mental model is reasonable and feasible.

TPTM (Task, Phase, Trajectory, Mentality) is the CBM proposed in this paper.

4 Freehand Tracking Algorithm Based on TPTM

4.1 Shape TPTM

In this section, we will show the way to shape the TPTM model. Cognitive tasks are described as a linear list, in which each item includes the name of the tasks, the related objects and their positions.

In order to build the TPTM, we requested the user to swear data glove and a position tracker on his hand. For each task, we asked her to practice the cognitive task with a natural style until she was skillful to it. Then, we asked her to perform this task R times while the data from the data glove and the position tracker was recorded.

Based on analysis of the data in the form of curves and of the videos observation, phases can be manually divided. A cognitive task corresponds to a series of phases. Each phase is described with ID number and entrance of trajectory.

A trajectory is a curve of a parameter of the hand model vector. The 'Trajectory' is described with trajectory parameters including curvature, mean value and mean square deviation.

Mentality includes cognitive goal, cognitive burden, attention, memory, training and so on. Mentality will be described with cognitive information.

The data structure of the shaped TPTM model is shown in Figure 5. 'T' and 'P' can be manually determined, 'T' is determined by learning, and 'M' is described with cognitive information.

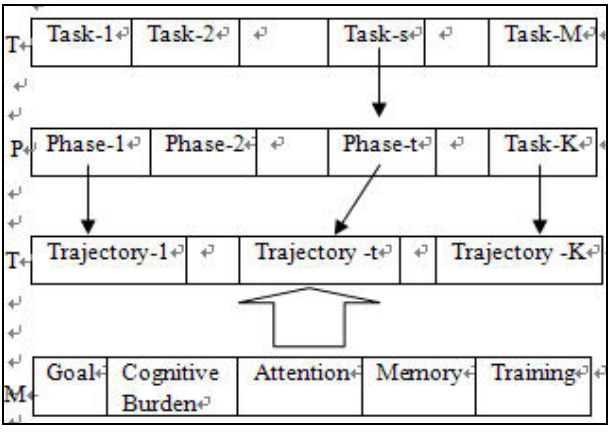


Fig. 5. The data structure of our TPTM model, which is organized as a tree structure. In the process of build CBM, we requested the user to achieve cognitive tasks with low cognitive burden and in the form of natural style. 'M' is described with cognitive information.

4.2 Cognitive Information Extraction

As a part of the mentality model, extraction and presentation of cognitive information is the key to distinguish different tasks and phases in TPTM model, so as that all cognitive tasks can be performed with correct way. We extracted a part of cognitive information, and some of them are shown in Fig. 6.

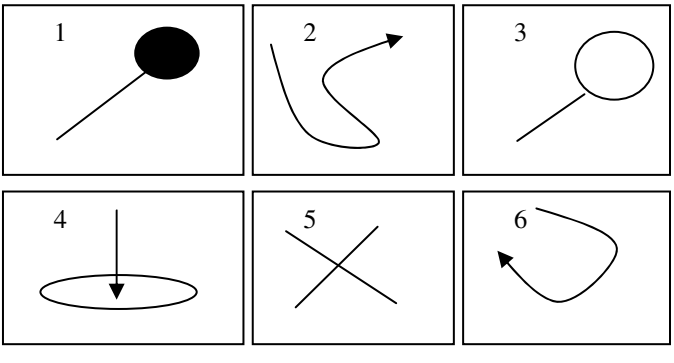


Fig. 6. Extracted basic cognitive information and its presentation. From left to right, Top: to grab, to move along the trajectory, to release; Bottom: to insert, to fail, to rotate.

For example, signal "1" means that the user should begin to grab an object in the step; signal "5" mans that the current phase is failed and the user should repeat the phase; signal "4" means that the user should insert the grabbed object into another.

Feedbacks are also used as important cognitive information. (1) if the user adjusts her hand with a style far from the previously learned model, computer will provide a feedback with a twinkling style on screen. (2) if the user fails to the current task, computer will warning signals on screen.

Cognitive information not only simplifies programming but also connects the user and computer.

4.3 Sample

According to our study, the number of sampling for a variable of the hand model is subjected to the following factors: curvature on the model curve at current frame, the Hausdorff distance of the tracked hand model at the previous frame and the mean square deviation on the curve at current frame. Firstly, on those points with larger curvature, the larger number of sampling should be drawn because of larger indeterminacy and unpredictability. Secondly, it is reasonable that if the better accuracy at current frame is achieved, the larger number of sampling should be used. Thirdly, the larger the mean square deviation means the larger indeterminacy.

Based upon the above analysis, the number of sampling for a variable of the hand model at a frame is computed in the following formula:

$$N = \alpha + \beta C + \gamma \ln H + \delta \ln D \tag{2}$$

where C, H and D are the curvature on the model curve at current frame, the Hausdorff distance of the tracked hand model at the previous frame, the mean square deviation on the curve at current frame, respectively, and α, β, γ and δ are empirical constants. The formula (2) is featured with combination of the history information and the current characteristics of TPTM model.

As to uniform distribution, the samples are drawn within area $[u-D, u+D]$, where u is the value on the model curve at current frame.

4.4 Freehand Tracking Algorithm

Based upon the TPTM model, our natural freehand tracking algorithm is designed as follows.

1. Build TPTM model for the user.

   (1) Make sure that the user understands the tasks in our VAS system.
   (2) With the help of data glove and position tracker on the user's hand, train the user to perform the task with natural style and low cognitive burden. After the user is skillful of the task, data from data glove and position tracker is recorded.
   (3) Compute model curve for each variable of hand model vector, the mean square deviation of each model curve, the curvature for each point on each model curve.
   (4) Construct a tree as shown in Figure 4.

2. Repeat
{
(1) Predict task and phase using cognitive information displayed on screen.
(2) Predict each variable according to corresponding model curve.
(3) Compute the number of samples for each variable using the formula (2).
(4) Samples are drawn by combining uniform distribution as well as Gaussian distribution.
(5) Estimate the 3D hand model using PF tracker.
(6) Output the 3D hand model.
} Until the cognitive tasks are performed.

The samples are composed of two parts. The one is particles from Gaussian distribution at the point with the biggest probability, and another one is particles from uniform distribution.

For simplicity, the above proposed algorithm is called TA-TPTM (Tracking Algorithm based on TPTM) in this paper. Different from PF tracker, TA-TPTM predicts the value at the next frame and draws particles based on the TPTM model, and the number of samples change over frames, decided by the curvature and the mean square deviation of model curve as well as the accuracy acquired at the last frame.

5 Experimental Results and Analysis

5.1 Experimental Settings

We use a color CCD calibrated camera ZT-QCO12 with a 4 mm lens that captures 640 × 480 video at 30 Hz. Our computer is Intel®, CoreTM, Quad CPU 2.66 GHz, 4G memory. We employ a 3D hand model with 26 degrees of freedom (DOF), 6 DOF for the global transformation and 4 DOF for each finger. The initialization of 3D hand model is completed based on cognitive behavior models.

5.2 The Results of Our Algorithm

Figure 7 show us some of the tracked freehand, it demonstrates that TA-TPTM can successfully track freehand which is implementing a cognitive task.

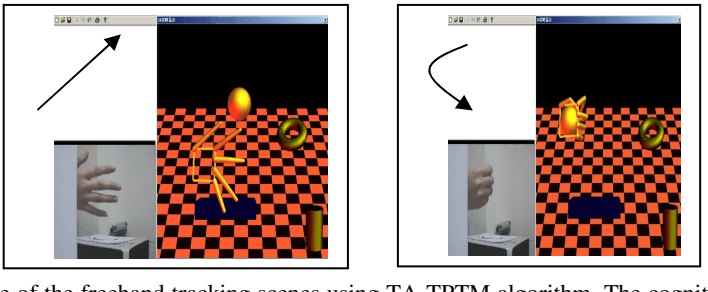


Fig. 7. Some of the freehand tracking scenes using TA-TPTM algorithm. The cognitive task is to move a 3D sphere from one place to another.

5.3 Reference Experiments

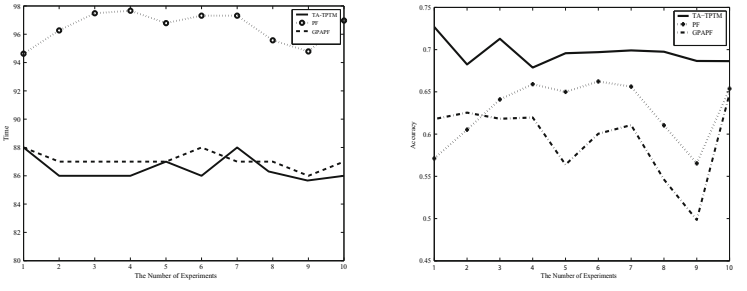
Comparison with GPAPF [27] and PF [CLZ05]: We test the reference algorithms, Gaussian Process Annealed Particle Filter (GPAPF) [27], proposed by Leonid Raskin et al. and PF [CLZ05] by Chen Rui et al., and compare the results with TA-TPTM. We compare these algorithms by means of the time cost and the accuracy for each frame. The accuracy of the tracking algorithms is measured by

$$A_m = e^{-\lambda H_m} \tag{3}$$

where H_{m} is Hausdorff distance between the projection of the 3D hand model onto the image plane and the image, of the method m, where m= *TA-TPTM, PF or GPAPF,* λ is an experiential const which is set as 0.01 in this experiment.

One of the key issues in evaluating system performance is the availability of ground-truth data. Obtaining ground truth data for 3D hand pose estimation is a difficult problem [3]. We project the hand model on the input image to show how well the projection match the image data, which can be measured by Hausdorff distance [14]. Feature extraction from image data is achieved by multiple scale approach [34].

Each algorithm of TA-TPTM as well as the reference algorithms was repeated 10 times, and the time cost and the accuracy are shown in Figure 8. The parameters in the formula (1) are as follows: α=2, β=1000, γ=0.1, δ=0.01. The number of particles drawn in TA-TPTM, PF and GPAPF is 5, 10 and 10 respectively.



(a) (b)

Fig. 8. The performance comparisons of TA-TPTM, PF and GPAPF. (a) the time cost. (b) the accuracy.

In Figure 8(a), among the three algorithms, TA-TPTM has the smallest time cost while PF has the largest time cost in each frame. In the 10 runs, the average time cost for each frames of TA-TPTM, PF and GPAPF is 81.6 ms, 96.4 ms and 86.7 ms respectively.

In Figure 8(b), TA-TPTM has the largest accuracy of the three algorithms in each frame of the 10 experiments. The average accuracy of TA-TPTM, PF and GPAPF is 0.696326, 0.627457 and 0.594864 respectively.

In order to comprehensively evaluate various algorithms, we further define the following evaluation criterion

$$Ratio_m = T_{average} \frac{A_m}{T_m} \tag{4}$$

where $T_{average}$ is the average time cost of the algorithms including the proposed algorithm and the reference algorithms, A_m and T_m are the average accuracy and the average time cost of the mth algorithm for per frame, respectively, where m =TA-TPTM, GPAPF or PF. According to formula (4), the larger the accuracy and the less the time cost, the better the comprehensive evaluation; if the time cost of a algorithm is equal to the average time cost of the all algorithms, the larger the accuracy, the better the comprehensive evaluation. Therefore, the proposed comprehensive evaluation is objective, fair and effective to evaluate an algorithm.

The ratios of the three methods are shown in Figure 9.

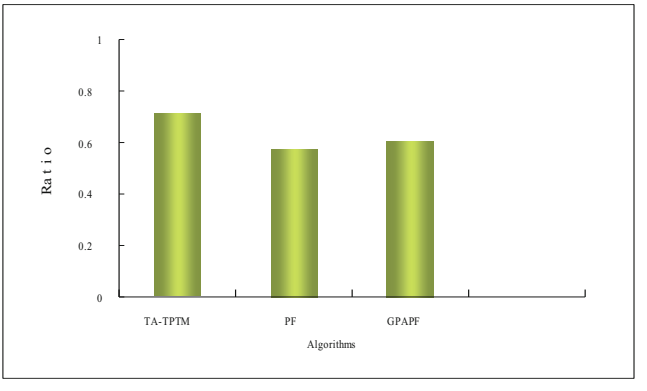
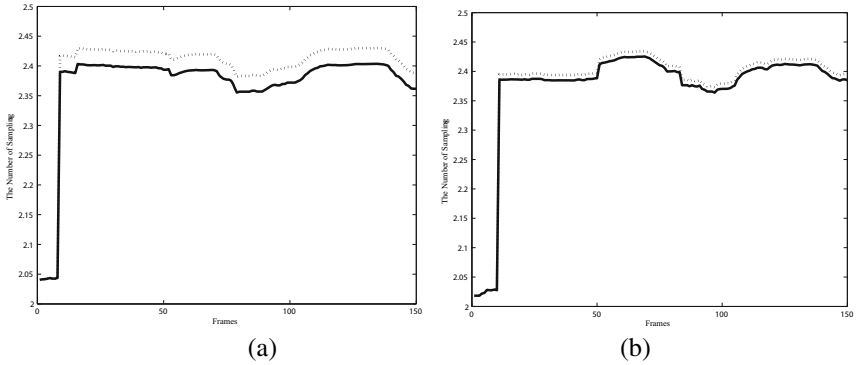


Fig. 9. The comprehensive evaluation, ratios, of the three methods. This evaluation takes the time cost and the accuracy into consideration. The ratio is defined in formula (3).

Figure 8 and Figure 9 show that compared with GPAPF or PF, the time cost, the accuracy and the comprehensive evaluation of TA-TPTM are improved.



(a) (b)

Fig. 10. The number of sampling for the proximal interphalangeal (PIP) in two experiments. (a) on index finger. (b) on middle finger.

For each variable in hand model, the number of sampling varies with frames, some of which are shown in Figure 10. The variable changes in the same way in the two experiments through their magnitudes are different. According to formula (2), the smaller the *C*, the *H* and the *D*, the smaller the N for a variable.

From Figure 10 we observed that the number of sampling dynamically changes with frames according to the history information and the TPTM model.

In order to further investigate the affection of the parameters *C*, *H* and *D* in the formula (2) on the performance of our algorithm, we experimented with $N=\alpha_1+\tau_1 C$, $\alpha_2+\tau_2 H$ *and* $\alpha_3+\tau_3 D$, the results are shown in Figure 11. Here, $\alpha_1=2$, $\tau_1=5000000$, $\alpha_2=1$, $\tau_2=0.08$, $\alpha_3=0.5$, $\tau_3=0.4$.

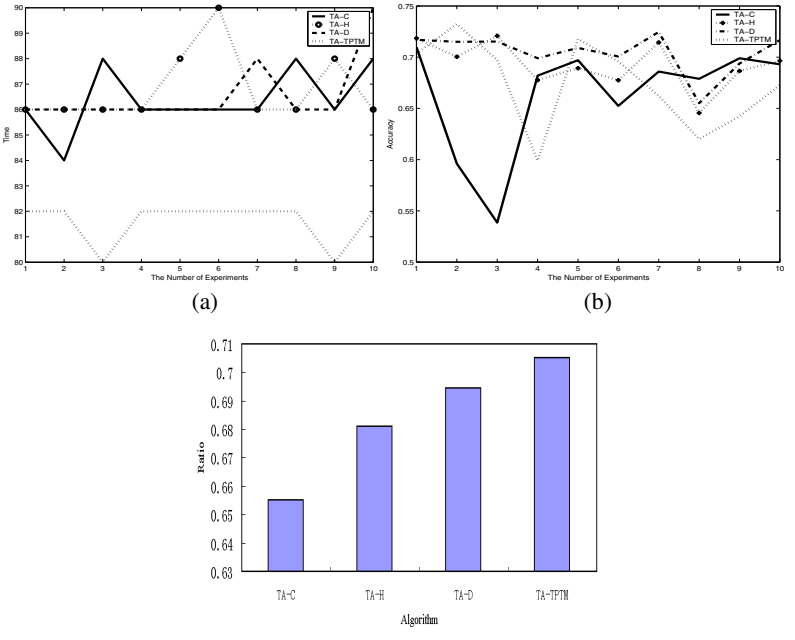


Fig. 11. The experimental results of our method with N=α1+τ1C, α2+τ2H and α3+τ3D corresponding to TA-C, TA-H and TA-D where C refers to the curvature, H refers to the Hausdorff distance, and D refers to the mean square deviation. (a) the average time cost. (b) the average accuracy. (c) the average comprehensive evaluation. Each experiment was repeated 10 times.

Figure 11 shows that on average, the time cost, the accuracy as well as the comprehensive evaluation are improved after all of the *C*, the *H* and the *D* are taken into consideration. The positive influence of the *C*, the *H* and the *D* on the time cost is the same. The *D* has the largest positive influence on the accuracy while the *C* has the smallest positive influence on the accuracy. The positive influence of the *D* on the comprehensive evaluation is the largest among the *C*, the *H* and the *D* while the *C* has the smallest influence.

Comparison with VLMM [23]: According to VLMM [23], distinct gestures are handled by using different models; transitions between them are learned with a variable-length Markov model that captures both high- and low-level structure. Much different from VLMM, TA-TPTM is featured with several points: (1) FTBM pays attention on users' cognitive and behavioral model, while VLMM pays little attention on users cognition. (2) TA-TPTM connects freehand tracking with cognitive tasks. (3) TA-TPTM switches between different model representations of hand posture by fusing probability transformation matrix for basic operations, cognitive information and image features, while TA-TPTM updates and automatically switches between different model representations of hand posture that correspond to distinct gestures by HMM. In 10 out of the 5000 frames, the performance of TA-TPTM and VLMM is shown in Fig. 12. We observed from Fig. 12 that TA-TPTM reduces time cost compared with VLMM.

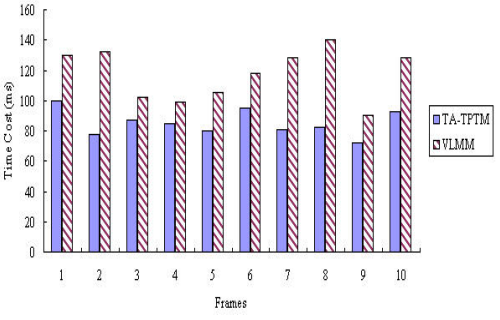


Fig. 12. The time cost comparison of TA-TPTM and VLMM

5.4 Applications

TA-TPTM has successfully been applied to virtual assembly system, it has also been applied several other practical systems by now.

Pointer Device. We built CBM for users in a pointer system, the freehand is tracked by TA-TPTM in real time (Fig. 13) and 3D hand is used as a 3D pointer.

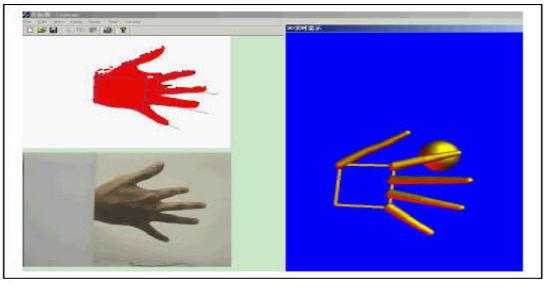


Fig. 13. The user is interacting with a 3D sphere using his freehand. Top-left: extracted cognitive information. Bottom-left: hand image from video. Right: reconstructed 3D hand model is touching a 3D sphere.

We also shaped CBM in a 3D network system, and the 3D network can be navigated with tracked 3D hand model (Fig. 14).

Fig. 14. The user is navigating with 3D network using his freehand. Left: experimental scene. Right: a scene of 3D network.

6 Conclusion and Future Work

Our experimental results show that the proposed approach raises the quality of each sampled particle or avoid sampling "poor" particles which appear with low probability in each frame, and it tracks 3D freehand in real-time with high accuracy. The number of the drawn particles is reduced up to 5 and the tracking speed is decreased down to 81 ms per frame. The proposed method is also valuable to many related issues, such as human body tracking and hand recognition as well as cognitive analysis in human-computer interaction.

Acknowledgments. This paper is supported by National Natural Science Foundation of China (No. 61173079, No.60973093, No. 61173078), Natural Science Foundation for Distinguished Youth Scholar of Shandong Province (No.JQ200820), Key Project of Natural Science Foundation of Shandong Province (ZR2011FZ003), NSFC project no 61173124, 61170318.

References

1. Agarwal, A., Triggs, B.: Tracking Articulated Motion Using a Mixture of Autoregressive Models. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004, Part III. LNCS, vol. 3023, pp. 54–65. Springer, Heidelberg (2004)
2. Anderson, J.R., Corbett, A.T., Koedinger, K.R., Pelletier, R.: Cognitive tutors: Lessons learned. J. Learn. Sci. 4, 167–207 (1995)
3. Erol, A., Bebis, G., Nicolescu, M., Boyle, R.D., Twombly, X.: A Review on Vision-Based Full DOF Hand Motion Estimation. In: IEEE Computer Society Conference on Computer Vision And Pattern Recognition, San Diego, CA, USA, pp. 15–22 (2005)
4. Bray, B., Koller-Meier, E., Muller, M., Van Gool, L., Schraudolph, N.N.: 3D Hand Tracking By Rapid Stochastic Gradient Descent Using A Skinning Model. In: 1st European Conference on Visual Media Production, London, pp. 231–237 (2004)

5. Bray, M., Koller-Meier, E., Van Gool, L.: Smart Particle Filtering for High-dimensional Tracking. Computer Vision and Image Understanding 106, 116–129 (2007)
6. von Hardenberg, C., Brard, F.: Bare-Hand Human-Computer Interaction. In: Proceedings of the ACM Workshop on Perceptive User Interfaces, pp. 1–8. ACM Press, New York (2001)
7. Cipolla, R., Ollinghurst, N.J.: Human Robot Interface by ointing with Uncalibrated Stereo Vision. Image and Vision Computing 14, 171–178 (1996)
8. Wang, C., Gao, W., Shan, S.: An approach based on phonemes to large vocabulary Chinese sign language recognition. In: Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 393–398. IEEE Press, New York (2002)
9. Deutscher, J., Blake, A., Reid, I.: Articulated Body Motion Capture By Annealed Particle Filtering. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1144–1149. IEEE Press, New York (2000)
10. Deutscher, J., Davison, A., Reid, I.: Automatic Partitioning of High Dimensional Search Spaces Associated With Articulated Body Motion Capture. In: Proceedings of Conference on Computer Vision and Pattern Recognition, pp. 187–193. IEEE Press, New York (2001)
11. Daubney, B., Gibson, D., Campbell, N.: Real-time Pose Estimation of Articulated Objects Using Low-level Motion. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE Press, New York (2008)
12. Erol, A., Bebis, G., Nicolescu, M., Boyle, R., Twombly, R.: Vision-based Hand Pose Estimation: A Review. Computer Vision and Image Understanding 108, 52–73 (2007)
13. Hix, D., Hartson, R.H.: Developing User Interfaces - Ensuring Usability Through Product and Process. John Wiley & Sons, New York (1993)
14. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing Images Using the Hausdorff Distance. IEEE Trans. Pattern Analysis and Machine Intelligence 15, 850–863 (1993)
15. Zhou, H., Huang, T.S.: Tracking Articulated Hand Motion with Eigen Dynamics Analysis. In: International Conference on Computer Vision, Nice, France, pp. 1102–1109 (2003)
16. John, B.E.: Cognitive modeling in human-computer interaction. In: Proceedings of Graphics Interface, pp. 161–167 (1998)
17. Kato, M., Chen, Y.W., Xu, G.: Articulated Hand Tracking by PCA-ICA Approach. In: Proceedings of IEEE Conference on Automatic Face and Gesture Recognition, pp. 329–334. IEEE Press, New York (2006)
18. Lenman, S., Bretzner, L., Thuresson, B.: Computer Vision Based Hand posture Interfaces for Human–Computer Interaction. Department of Numerical Analysis and Computer Science, Sweden (2002)
19. Lee, M.S., Weinshall, D., Cohen Solal, E., Colmenarez, A., Lyons, D.: Computer Vision System for On-screen Item Selection by Finger Pointing. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 329–334. IEEE Press, New York (2001)
20. Lin, J., Wu, Y., Huang, T.S.: Capturing Human Hand Motion In Image Sequences. In: Workshop on Motion and Video Computing, pp. 99–104 (2002)
21. McAllister, G.: Hand tracking for behaviour understanding. Image and Vision Computing 20, 827–840 (2002)
22. Mental models and usability, `http://www.lauradove.Info/reports/mentalTechnicalreport`

23. Stefanov, N., Galata, A., Hubbold, R.: A real-time hand tracker using variable-length Markov models of behaviour. Computer Vision and Image Understanding 108, 98–115 (2007)
24. Oka, K., Sato, Y., Koike, H.: Real-time tracking of multiple fingertips and gesture recognition for augmented desk interface systems. In: Fifth IEEE International Conference on In Automatic Face and Gesture Recognition, pp. 411–416. IEEE Press, New York (2002)
25. Chen, Q., Emil Petriu, M., Nicolas Georganas, D.: 3D Hand Tracking and Motion Analysis with a Combination Approach of Statistical and Syntactic Analysis. In: IEEE International Workshop on Haptic Audio Visual Environments and their Applications, pp. 56–61. IEEE Press, New York (2007)
26. Wang, R.Y., Popović, J.: Real-Time Hand-Tracking with a Color Glove. ACM Transactions on Graphics 28, 1–8 (2009)
27. Raskin, R.L., Rudzsky, E.: Dimensionality Reduction for Articulated Body Tracking. In: 3DTV Conference, Kos Island, pp. 1–4 (2007)
28. Stenger, B.: The Grid: Model-Based Hand Tracking Using A Hierarchical Bayesian Filter. PhD Thesis, Department of Engineering, University of Cambridge (2004)
29. Lin, T., Zha, H., Lee, S.U.: Riemannian Manifold Learning for Nonlinear Dimensionality Reduction. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 44–55. Springer, Heidelberg (2006)
30. Vaswani, N.: Particle Filtering For Large Dimensional State Spaces with Multimodal Observation Likelihoods. IEEE Transactions on Signal Processing 56, 4583–4597 (2008)
31. Wu, Y., Lin, J.Y., Huang, T.S.: Capturing Freehand Articulation. In: IEEE International Conference On Computer Vision, Vancouver, Canada, vol. 2, pp. 426–432 (2001)
32. Xu, X., Li, B.: Rao-Blackwellised: Particle Filter For Tracking with Application in Visual Surveillance. In: The 2nd Joint IEEE International Workshop, pp. 17–24 (2005)
33. Wu, X., Liang, W., Jia, Y.: Tracking articulated objects by learning intrinsic structure of motion. Pattern Recognition Letters 30, 267–274 (2009)
34. Feng, Z., Yang, B., Zheng, Y.: Research on features extraction from frame image. sequences. In: International Symposium on Computer Science and Computational Technology, pp. 762–766 (2008)

SPIDAR-Pen: A 2T1R Pen-Based Interface with Co-located Haptic-Visual Display

Liping Lin[1], Yongtian Wang[1], Katsuhito Akahane[2], and Makoto Sato[2]

[1] School of Computer Science and Technology, Beijing Institute of Technology, Beijing, 100081, China
[2] Precision and Intelligence Laboratory, Tokyo Institute of Technology, Yokohama, 226-8503, Japan
{linliping,wyt}@bit.edu.cn,
kakahane@hi.pi.titech.ac.jp, msato@pi.titech.ac.jp

Abstract. This paper proposes a string-driven pen-based interface, which is capable of two translational and one rotational (2T1R) interaction, and co-located haptic-visual display. The inputs of interface are translation and rotation of pen tip tracked by strings. Simultaneously, force and moment feedbacks are displayed at pen tip via string tensions. In addition, visual display and haptic display are coincident and well integrated. With this device, user can create, design and manipulate virtual world with force feedback by drawing, selecting, moving or rotating objects on display in a natural and intuitive way. Experiments show that this interface can greatly improve manipulation efficiency. And a system which combines this interface and simulated 2D physical world for physics education is developed to demonstrate its merit.

Keywords: Pen-based, haptic interaction, multimodal, virtual reality.

1 Introduction

There is a growing trend towards multimodal system in the design of human-computer interfaces [1]. Because multimodal system is a more natural and effective interface that combines different input modes such as speech, pen, touch, hand gestures and provides multiple output modalities, i.e. visual, auditive and tactile, etc, in a coordinated manner [2].

As for input mode, pen-based interface is becoming more and more popular, owing to its adaption to the natural interaction skill that users employ in everyday life. Moreover, it promises more accurate manipulation. Therefore, it is applied in various areas, ranging from PDA (Personal Digital Assistants) [3], to traditional PC[4] and electronic whiteboards[5-6].

However, in most existing pen-based interaction, pen is operated in point-interaction style for selecting, moving, dragging objects or writing text. And pen rolling around its axis has rarely been considered as an input modality. But in some applications, rotational input also is helpful or necessary for effective manipulation, such as virtual playland for children [7], or some 2D games like puzzle. Xiaojun Bi et

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 166–177, 2012.
© Springer-Verlag Berlin Heidelberg 2012

al [8] conducted an exploration of pen rolling for pen-based interaction and demonstrated that pen-rolling interaction is beneficial in three aspects: enhanced stimulus-response compatibility in rotation tasks, multi-parameter input and simplified mode selection. They also proposed a system to determine the rolling angle of pen. That system consists of an 8-camera Vicon motion tracking system and four passive reflective markers placed on pen.

As for output interface, one important modality is force feedback. The sensing of forces is closely coupled to both the visual system and one's spatial sense; the eyes and hands work in concert to explore and manipulate objects. Such sensorial information offers the potentiality of enhancing the manipulative interaction capabilities in virtual environments (VEs) for activities. Furthermore, some researchers have demonstrated that haptic information is useful for reducing the perceived musculoskeletal loading as measured through pain and discomfort in completing the task [9].

In 2D interaction field, several researchers have attempted to integrate haptic interfaces and computer scene for establishing multimodal systems. Noma et al [10] proposed a digital desk named "Proactive Desk" which can generates an omnidirectional translational force on a user's hand without any mechanical link or wire. Jongwon Back et al [11] designed a 2D haptic interface of 2-DOF which can be used for various purposes, and implemented a virtual air-hockey system by combining it with 3D image system. Solis, J et al [12] proposed a system named "Haptic Desktop". This interface provides co-located perceptual information (visual and haptic) and moreover, the conventional computer input components (mouse and keyboard) are replaced by 2-DOF haptic interface. And Solis, J et al also pointed out that the concept of co-location has becomes an important criterion for the design of multimodal interfaces. The term co-location of haptic and visual display means that the user can visually perceive an object at the same position in space as the haptic simulation, and that can reduce mental load of users and integrate information of different channels in an optimal way.

In this paper, a pen-based interface SPIDAR-Pen is presented. It performs in an interactive mode of two translational and one rotational (2T1R) interaction. Tracking the translation and rotation of pen tip and displaying force and moment are realized by a 2D string-based haptic device SPIDAR [13]. In addition, haptic display is calibrated with visual display such that the haptic and visual co-ordinate systems are coincident. Furthermore, it has some other advantages, such as low cost, scalability, etc. This paper is organized as follows: In Section 2, we present a brief description of SPIDAR-Pen interface. In Section 3, the implementation of the system is described. Section 4 shows the performance evaluation and an application demo of this interface in physics education. And Section 5 is conclusion.

2 The SPIDAR-Pen Interface

SPIDAR-Pen is a pen-based interface, as shown in figure 1. The pen tip is tied with four strings that driven by four motors respectively. The strings are always in tension. This interface has two main characteristics.

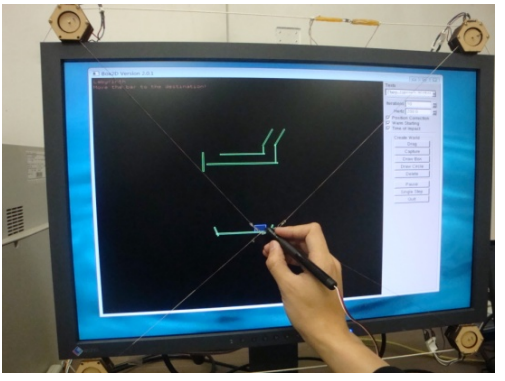


Fig. 1. The SPIDAR-Pen interface

(1) 3 DOF input and output. When user holds the pen and moves or rotates it on the screen surface, the translations of tip in horizontal and vertical directions and the rotation around its axis will be calculated via string lengths. The maximum rotational angle range is (-180, 180) degree. And also the translational forces and rotational moment can be displayed through string tensions.
(2) Haptic display and visual display are co-located. The pen is registered with visual display, so haptic and visual co-ordinate systems are coincident. Therefore, when user visually perceives that pen tip collides with other objects, the force feedback will be simulated at the same position. In addition, the pen is driven by strings which keeps the working space transparent and does not affect the visual display.

In a word, with this interface, user can design and manipulate virtual world by drawing, selecting, moving or rotating objects on display in a natural and intuitive way.

3 System Implementation

3.1 Hardware System

The design of the hardware system is characterized by its scalability, flexibility and low cost. Therefore, this system can be combined with common personal computers, adapted to panel displays of different sizes, and afforded by common users.

The system consists of four parts, as shown in figure 2.

(1). Controller

The processing chip is PIC24FJ256GB106, which is a 16 bit microcontroller chip. Its CPU speed is up to 16MIPS, and USB 2.0 module can work in full-speed mode. Its performance provides sufficient capability for controlling 4 motors and processing optical encoder pulses of 4 channels. This chip is cheap but competent for this application.

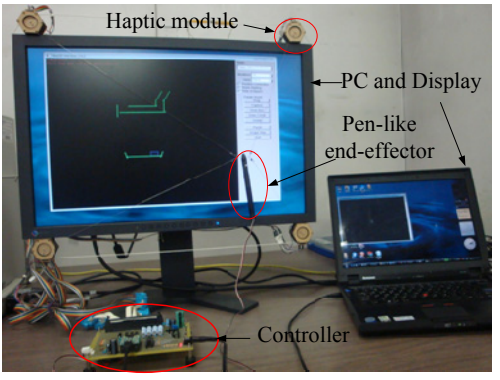


Fig. 2. Structure of SPIDAR-Pen system

(2). Haptic module
Haptic module consists of Mabuchi RE280 DC motor, pulley, optical encoder and string. Four modules are fixed at four corners of display. The shape of modules is hexagonal, which can be fit for different kind of display. And the maximum continuous force is 2.0N.

(3). Pen-like end-effector
As shown in figure 3, end-effector has special structure. A small stick is fixed through the tip, and then strings of haptic modules are attached at the stick. At the holding part of the pen, there are two programmable buttons, which can perform the functions of right and left buttons of common mouse.

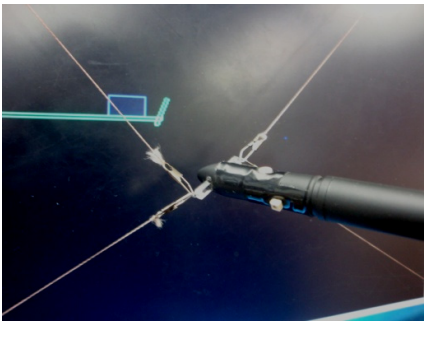


Fig. 3. Pen-like end-effector

(4). Computer and display
Common computer and display are sufficient for the requirements of this system. Furthermore, it also can work with large panel displays. This demonstrates the scalability and flexibility of this device.

3.2 Position and Orientation Calculation Algorithm

As shown in figure 3, there is a small stick attached at the tip of end-effector. In order to calculate the position and orientation of end-effector tip, the positions of two end-points of the stick should be calculated firstly.

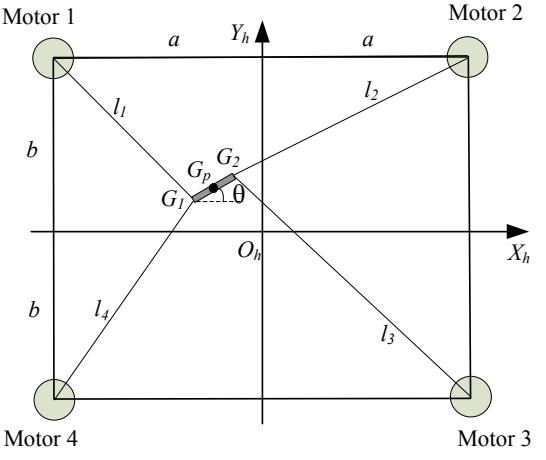


Fig. 4. Calculation of end-effector tip position and rotation

Figure 4 shows that four DC motors are fixed at M_1 *(-a, b),* M_2 *(a, b),* M_3 *(a, -b),* M_4 *(-a, -b)* in $X_hO_hY_h$ coordinates system. Lengths of four strings are l_1, l_2, l_3, l_4 respectively. $G_1(X_1, Y_1)$ and $G_2(X_2, Y_2)$ are two endpoints of the small stick. Their positions can be obtained from Eq.1 and Eq.2.

$$\begin{cases} (X_1-(-a))^2+(Y_1-b)^2=l_1{}^2 \\ (X_1-(-a))^2+(Y_1-(-b))^2=l_4{}^2 \end{cases}. \tag{1}$$

$$\begin{cases} (X_2-a)^2+(Y_2-b)^2=l_2{}^2 \\ (X_2-a)^2+(Y_2-(-b))^2=l_3{}^2 \end{cases}. \tag{2}$$

The values of their positions are

$$\begin{cases} X_1=\dfrac{\sqrt{8b^2(l_1{}^2+l_4{}^2-2b^2)-(l_4{}^2-l_1{}^2)^2}}{4b}-a \\ Y_1=\dfrac{l_4{}^2-l_1{}^2}{4b} \end{cases}. \tag{3}$$

$$\begin{cases} X_2=-\dfrac{\sqrt{8b^2(l_2{}^2+l_3{}^2-2b^2)-(l_3{}^2-l_2{}^2)^2}}{4b}+a \\ Y_2=\dfrac{l_3{}^2-l_2{}^2}{4b} \end{cases}. \tag{4}$$

Therefore, the position of end-effector tip $G_p(X_p, Y_p)$ is

$$\begin{cases} X_p=\dfrac{X_1+X_2}{2} \\ Y_p=\dfrac{Y_1+Y_2}{2} \end{cases}. \tag{5}$$

And the rotation angle is

$$\theta = Arc\tan(\frac{Y_2 - Y_1}{X_2 - X_1}) \quad . \tag{6}$$

3.3 Tension Calculation Algorithm

The force display is 3 DOF. They are translational forces f_x , f_y in X direction, Y direction respectively and rotational moment m. And the displayed force is the resultant force of strings tensions $\tau_1, \tau_2, \tau_3, \tau_4$ as shown in Figure 5. φ_1, φ_2, φ_3 and φ_4 are directional vector of four strings, which can be calculated from the positions of four motors and two endpoints of small stick, G_1 and G_2. They also can be expressed as $(\varphi_{ix}, \varphi_{iy})(i = 1,2,3,4)$.

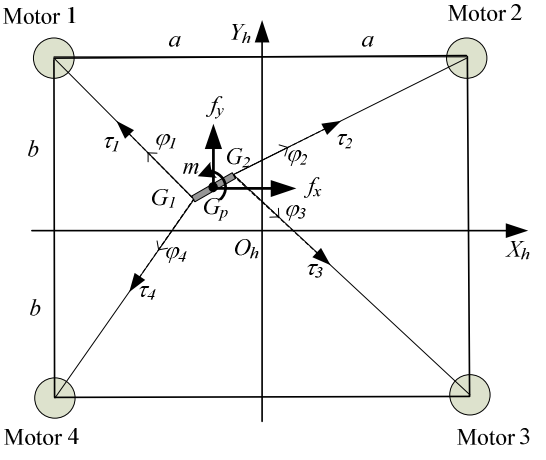


Fig. 5. Analysis of displayed force and strings tensions

Therefore, the displayed force is

$$\boldsymbol{F} = \begin{bmatrix} f_x \\ f_y \\ m \end{bmatrix} = \begin{bmatrix} \varphi_{1x} & \varphi_{2x} & \varphi_{3x} & \varphi_{4x} \\ \varphi_{1y} & \varphi_{2y} & \varphi_{3y} & \varphi_{4y} \\ X_1\varphi_{1y} - Y_1\varphi_{1x} & X_2\varphi_{2y} - Y_2\varphi_{2x} & X_2\varphi_{3y} - Y_2\varphi_{3x} & X_1\varphi_{4y} - Y_1\varphi_{4x} \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{bmatrix} \tag{7}$$

This can be rewritten as

$$\boldsymbol{F} = \boldsymbol{W}\boldsymbol{\tau} \quad . \tag{8}$$

The values of (X_1, Y_1) and (X_2, Y_2) in Eq. (7) can be attained from Eq. (3), Eq. (4). And the tension of strings should be limited to $[\tau_{\min}, \tau_{\max}]$. The minimum tension is $\tau_{\min}$, which maintain the string always in tension, the maximum tension is $\tau_{\max}$, which is limited by motor output torque.

Eq.(8) can't be solved directly, so it is transformed into optimization problem as follows. [14]

$$J = \left\| \boldsymbol{W}\boldsymbol{\tau} - \boldsymbol{F} \right\|^2 + \lambda \left\| \boldsymbol{\tau} \right\|^2 \to \min \quad (\tau_{\min} \le \tau_i \le \tau_{\max}) \tag{9}$$

In Eq.(9), the first item means that the difference between the displayed force and the desired force should be as small as possible. And parameter λ represents the balance of force display accuracy and stability. Then the optimization problem can be solved as quadratic programming problem efficiently.

3.4 Registration of Haptic Pen

Registration of SPIDAR-Pen is implemented to assure the coincidence between the haptic and graphical information. In figure 6, the width and height of display are D_x, D_y respectively, and its resolution is $D_{px} \times D_{py}$. The position of the origin point O_v of visual coordinates system $X_vO_vY_v$ is (x_{v0}, y_{v0}) in haptic coordinates system $X_hO_hY_h$. The position of end-effector tip (X_p, Y_p) in $X_hO_hY_h$ can be attained by Eq.5. Then this position values should be transformed to values (X_v, Y_v) in visual coordinates system $X_vO_vY_v$ according to the following equations.

$$\begin{aligned} X_v &= (X_p - x_{v0}) \frac{D_{px}}{D_x} \\ Y_v &= -(Y_p - y_{v0}) \frac{D_{py}}{D_y} \end{aligned} \tag{10}$$

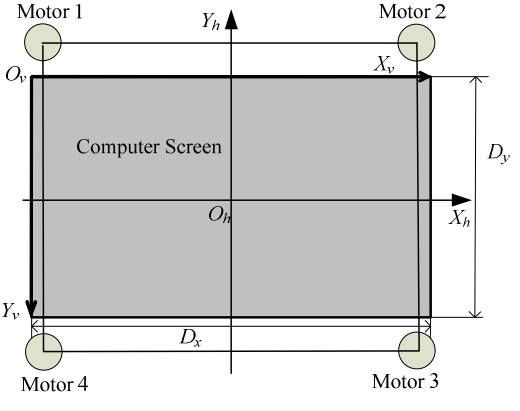


Fig. 6. Registration of SPIDAR-Pen with display

If the origin of haptic coordinates system $X_hO_hY_h$ is at the center of computer screen, then

$$x_{v0} = -\frac{D_x}{2} \quad . \qquad (11)$$
$$y_{v0} = \frac{D_y}{2}$$

In this system, the mouse functions are replaced by pen-like end-effector. And the movement of cursor is driven by the movement of end-effector tip using the mouse Win32 APIs. Theoretically, cursor position should be completely coincident with the position of pen tip on display after calibration. However, there is deviation between the positions of cursor and pen tip, which is caused by the errors of string lengths and motor positions. An experiment was carried out to demonstrate the registration accuracy and its result is shown in figure 7. 81red points are distributed on display in array. During the experiment procedure, pen tip is fixed at these red points one by one, and at the same time, the positions of cursor are captured, which are shown in blue points. It can be found that the deviation is small. However, when pen tip is near to the edge of display, the deviation will grow larger.

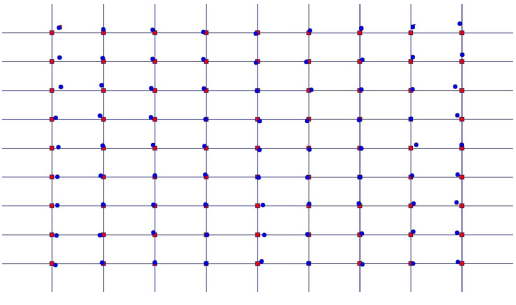


Fig. 7. Distribution of pen tip positions (red points) and cursor positions (blue points) on display

4 Application and Evaluation

4.1 Experiments and Evaluation

Experiments are conducted to evaluate the manipulation efficiency of this interface. Manipulation tasks involve moving and rotating objects, which are common operations in designing or manipulating virtual world. The completion time is compared to the time for accomplishing the same tasks using common mouse. By means of mouse manipulation, rotating objects is usually realized by two steps: to fix the rotation center firstly and then to drag a corner of the object to rotate it (e.g., as in Phun).

The experiments employ two manipulation tasks. One is moving and rotating a peg to fit a hole, as shown in figure 8(a). Another one is moving and rotating a box to pass through an irregular path, as shown in figure 8(b). 6 subjects of 22-28 ages accomplished these tasks in two interaction methods by using common mouse and

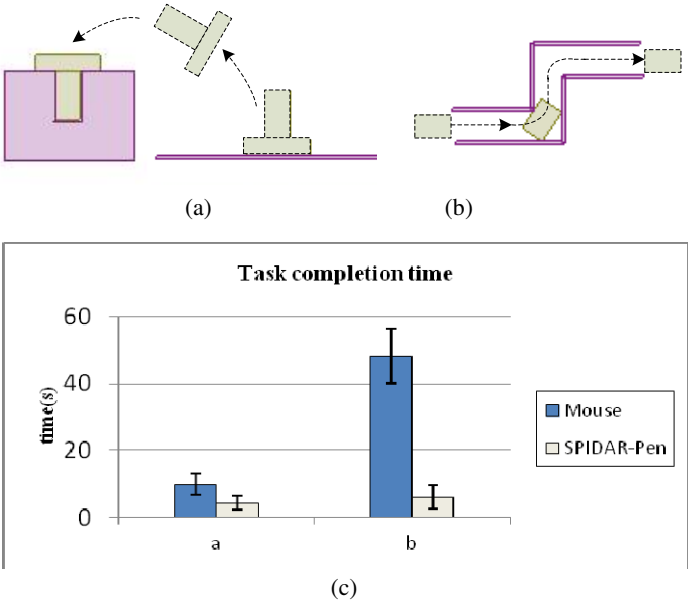


Fig. 8. (a) Peg-in-hole task. (b) Moving and rotating an object throughout an irregular path. (c) Comparing completion time of manipulations by mouse and SPIDAR-Pen respectively.

SPIDAR-Pen respectively. Before experiments, subjects had sufficient trials to familiarize themselves with the tasks.

The complement time is compared in figure 8(c). It shows that manipulation via SPIDAR-Pen owns higher efficiency. But in experiment (a), there is no great efficiency difference between these two manipulation methods. This is because the task is simple. Although the task should be completed by fixing, rotating and moving object step by step, the conversion from one operating mode to another is easy and quick. However, the time difference becomes apparent in experiment (b). The reason is that this task requires high coordination between motions of moving and rotating object, while this is difficult to be implemented by manipulating mouse.

4.2 Application in Physics Education

SPIDAR-Pen can be applied in various areas, and one promising application is for interaction with 2D simulated physical world. For example, it would become very fascinating if it is combined with Phun, which is an educational and entertainment software for designing and exploring 2D multi-physics simulation in a cartoony fashion [15]. In our research, a pen-based interaction system for physics education with co-located planar haptic and visual feedback is realized. This system combines SPIDAR-Pen and simulated 2D physical world created by physical engine Box2D and 2D game engine HGE. With this system, user can create, design and manipulate simulated physical world by drawing, selecting, moving or rotating objects on computer display with force feedback in a natural and intuitive way.

Box2D is a free open source 2D simulator engine and performs constrained rigid body simulation. This engine also applies gravity, friction, restitution and so on. However, the Box2D physical world is admittance environment, namely, the input is force ($\boldsymbol{F}_e$) applied at virtual object and the output is the position and posture ($\boldsymbol{V}_e$) of the virtual object. But the haptic device is an impedance interface. The input is position and posture information ($\boldsymbol{V}_h$) of end-effector tip and the desired force ($\boldsymbol{F}_h$) is displayed. Therefore, these two parts can't be combined directly. Instead, virtual coupling is used to combine these two parts [16], as shown in figure 9. In addition, virtual coupling also performs as an impedance filter, which limits the maximum or minimum impedance presented by the haptic device in such a way as to guarantee the stability of haptic display [17].

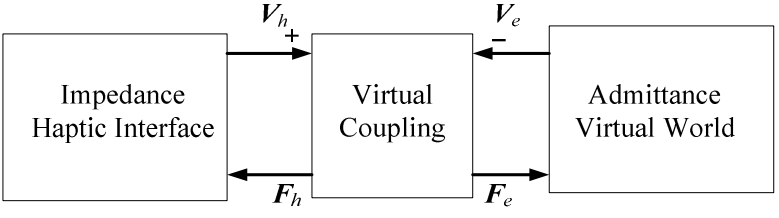


Fig. 9. Combination of haptic display with physical world

Fig.10 shows designing and operation interface of this system. The acceleration of gravity of this virtual physical world can be adjusted by slider. And user can create new objects like circle, box, and polygon in the world and can drag, move, rotate the existing objects with force feedback. The properties of objects, such as density, friction, damping, and restitution can be adjusted by sliders. In addition, the simulation can be started or stopped by clicking the button at right top corner.

This system provides an entertaining and cartoony tool for designing and carrying out physics experiments about friction, inertia, collision in various situations, and will greatly promote students' learning interest and creativity.

Fig. 10. Designing and operation interface for physics experiments

5 Conclusion

This paper proposed a novel string-driven pen-based interface. The inputs include not only the position of pen tip but also its rotation, which enables user to directly move and rotate virtual objects simultaneously. Correspondingly, there are 3 degree-of-freedom force displays, which are two translational forces and one rotational moment. In addition, the pen is registered with computer display, so user can manipulate virtual world in a natural and intuitive way. Experiments show that interaction via SPIDAR-Pen can greatly improve manipulation efficiency. This interface can be applied in various areas, such as physics education, animation, 2D games and CAD, etc.

Acknowledgments. This work was supported by the National High Technology Research and Development Program of China (863 Program). Grant No.:2009AA012106, 20080606063 and the Innovation Team Development Program of the Chinese Ministry of Education , Grant No.:IRT0606.

References

1. Tan, H.Z., Gray, R., Young, J.J., Irawan, P.: Haptic Cueing of a Visual Change-Detection Task: Implications for Multimodal Interfaces. In: 9th International Conference on Human-Computer Interaction, pp. 678–682. Lawrence Erlbaum Associates, New Jersey (2001)
2. Interactive Multimodal Systems, `http://www.hcsnet.edu.au/node/1126`
3. Long, A.C., Landay, J.A., Rowe, L.S.: PDA and Gesture Use in Practice: Insights for Designers of Pen-Based User Interfaces. Technical report, UCB//CSD-97-976,U.C. Berkeley (1997)
4. Li, Y., Guan, Z., Dai, G., Ren, X., Han, Y.: A Context-Aware Infrastructure for Supporting Applications with Pen-Based Interaction. Journal of Computer Science and Technology 18(3), 343–353 (2003)
5. Moran, T.P., Chiu, P., Van Melle, W.: Pen-based Interaction Techniques for Organizing Material on an Electronic Whiteboard. In: 10th Annual ACM Symposium on User Interface Software and Technology, pp. 45–54. ACM Press, Alberta (1997)
6. Pimentel, M.d.G., Prazeres, C., Ribas, H.: Documenting the Pen-Based Interaction. In: 11th Brazilian Symposium on Multimedia and the Web, pp. 1–8. ACM Press, Brazil (2005)
7. Dai, G.: Pen-based User Interface. In: 8th International Conference on Computer Supported Cooperative Work in Design, pp. 32–36. IEEE Computer Society, Xiamen (2004)
8. Bi, X., Moscovih, T., Ramos, G.: An Exploration of Pen Rolling for Pen-based Interaction. In: 21st Annual ACM Symposium on User Interface Software and Technology, pp. 191–200. ACM Press, Monterey (2008)
9. Dennerlein, J.T., Yang, M.C.: Haptic Force Feedback Devices for the Office Computer: Performance and Musculoskeletal Loading Issues. Human Factors 43(2), 278–286 (2001)
10. Noma, H., Yanagida, Y., Tetsutani, N.: The Proactive Desk: A New Force Display System for a Digital Desk Using a 2-DOF Linear Induction Motor. In: IEEE Virtual Reality 2003, pp. 217–224. IEEE Computer Society, Los Angeles (2003)

11. Back, J., Kang, J., Yong, H.-J., et al.: Design and Fabrication of a 2D Haptic Interface Device and the Realization of a Virtual Air-Hockey System. In: EuroHaptics 2006, pp. 613–616. EuroHaptics Society, Paris (2006)
12. Solis, J., Marcheschi, S., Portillo, O., et al.: The Haptic Desktop a Novel 2D Multimodal Device. In: 2004 IEEE International Workshop on Robot and Human Interactive Communication, pp. 521–526. IEEE Computer Society, Okayama (2004)
13. Sato, M.: Development of String-Based Force Display: SPIDAR. In: 8th International Conference on Virtual Systems and Multi Media, pp. 1034–1039. International Society VSMM, Gyeongju (2002)
14. Hasegawa, S., Inoue, M., Kim, S.: Tension Calculation Method for Tension Driven Force Display. Journal of the Robotics Society of Japan 22(5), 610–615 (2004)
15. Phun-2D Physics Sandbox, http://www.phunland.com/wiki/Home
16. Adams, R.J., Hannaford, B.: Control Law Design for Haptic Interfaces to Virtual Reality. IEEE Transactions on Control Systems Technology 10(1), 1–12 (2002)
17. Colgate, J.E., Stanley, M.C., Brown, J.M.: Issues in the Haptic Display of Tool Use. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 140–145. IEEE Computer Society, Pennsylvania (1995)

Generation of IFS Fractal Images Based on Hidden Markov Model*

Liliang Zhang[1] and Zhigeng Pan[2,3]

[1] Department of Computer and Information Engineering, Ningde Normal University, Ningde 352100, China
[2] Digital Media and HCI Research Center, Hangzhou Normal University, China
[3] School of Education Science, Nanjing Normal University, China
13859603598@139.com

Abstract. This paper presents a method for generation iterated function systems fractal attractor images based on hidden Markov model. The method can draw the shape and color of fractal images by using probability matrix. Furthermore, the paper also gives a method to show how to draw the local shape and color with multi-level by resolving the affine transformations of IFS into many affine transformations of sub-images. Finally, the effect of the method is showed by computer experiments in the simulation of the trees etc.

Keywords: Markov, hidden Markov model, iterated function systems, fractal.

1 Introduction

Iterated Function Systems is a typical method of fractal image generation. Its basic idea is that the global and local of the geometry object have the similar structure in the affine transformation. By selecting some affine transformation, the model is able to generate fractal attractor [1, 2, 3]. Resent years, IFS has become an important branch of the fractal studies. The theory and method were researched more in-depth and widely used in many fields [4, 5, 6]. In the research of simulating natural scenery with IFS attractor, Sprot [7] used the random parameter to control IFS, to generate IFS attractor automatically in certain parameters interval. In order to get the balance of organized and chaotic, Karel and Simant [8] introduced MRFS (Mutually Recursive Function Systems) to strengthen the control of attractor image. Martyn [9] through doing standardized treatment with IFS to achieve good deformation effect by changing the shape of attractor. Chang [10] came up with HFPS algorithm, by doing all kinds of affine transformation to get new fractal with a known fractal sets. In the research of iterative function system, Dekking [11] introduced the Markov theory, and putted forward the IFS system with Markov characteristics. In addition, Barnsely, John and Hart et al. [12, 13, 14, 15] did the further research with this problem, overcame the

* This research was supported by the Fujian natural sciences Foundation under Grant No.2011J01358.

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 178–187, 2012.
© Springer-Verlag Berlin Heidelberg 2012

limitations of the traditional IFS method, got the general iteration function system of more universal significance (MIFS), and carry out many applied research.

MIFS methods only consider the shape model of the fractal attractor and lack of color expression elements. If the shape and color of fractal images takes as a observable external appearance, it was determined by a series of hidden unobservable states. So we can use a hidden Markov model to describe this process. This article proposes an iterative function system based on hidden Markov model (HMIFS). By selecting probability distribution freely, the shape and color of the fractal images are controlled by Markov chain. It not only changes the shape of HMIFS fractal attractor, but also changes the color of HMIFS fractal attractor. Furthermore, the IFS affine transformation is resolved into many affine transformations in multi-levels. It can not only control the shape and color of the global structure, but can also control the shape and color of the local structure, which can draw rich and colorful fractal images .

2 Background Knowledge

2.1 MIFS Systems

An iterated function system consists of a complete metric space (X, d) together with a finite set of contraction mappings $\omega_i : \mathrm{X} \to \mathrm{X}$. The abbreviation "IFS" is used for "iterated function system". The notation for the IFS just announced is $\{\mathrm{X}; \omega_1, \omega_2, \cdots, \omega_N\}$ and its each contraction mapping with a probability is $p_i > 0, \sum_{i=1}^{N} p_i = 1$. The contraction mappings often take the following form of affine transformations:

$$\omega_i \begin{pmatrix} x \\ y \end{pmatrix} = A_i X + B_i, i = 1,2,\cdots,N \tag{1}$$

If the contraction mappings are selected by Markov chain, they construct an iterated function system with Markov characteristics. The probability of next a second mapping and former a mapping closely related in the iterative calculation process. Let:

$$p_{ij} = P\{x_n = j \mid x_{n-1} = i\} \ (n > 0) \tag{2}$$

Where p_{ij} is conditional probability, then the matrix $P = (p_{ij})_{N \times N}$ is called Markov transition probability matrix, and the transition probability is $p_{ij} \geq 0, \sum_{j=1}^{N} p_{ij} = 1, i = 1,2,\cdots,N$, where p_{ij} is the probability of mappings ω_j when contraction mapping ω_i was chosen last time. At this time,

$\{X; \omega_1, \omega_2, \cdots, \omega_N, P\}$ is defined a Markov iterated function system (MIFS). MIFS can generate more complex attractor images than IFS by using Markov transition probability matrix with zero elements. Therefore, it expanded the range of IFS fractal image. If the different transition probability matrix is used then we can get the different fractal images of MIFS.

2.2 Hidden Markov Model

Hidden Markov model is a mathematical model which based on the basis of Markov process. It consists of two random process which is related. One is the hidden Markov chain (can't be observed) and has a finite state, it's $\{ S_n : n = 1, 2, \cdots, N \}$. The other is a random process which related to each state of Markov chain and can be observed, it's $\{ O_m : m = 1, 2, \cdots, M \}$.

Each random process with a group of probability distributions. Where conditional probability $P(O_m = v_k \mid S_n = i_n, O_{m-1} = v_{k_{m-1}}, \cdots, O_1 = v_{k_1}, S_1 = i_1)$ is known (we assume the conditions probability is have nothing to do and the value of $O_{m-1}, \cdots, O_1$). Where $\{ S_n : n = 1, 2, \cdots, N \}$ is called hidden Markov chain, $\{ O_m : m = 1, 2, \cdots, M \}$ is called observation chain. The abbreviation "HMM" is used for "hidden Markov model. A HMM is confirmed by the probability distribution of observation value, the initial state distribution and the transition probability distribution of Markov chain and Markov chain state is known. HMM has widely applied, and has been applied successfully in pattern recognition, bioinformatics and image processing etc [16, 17, 18, 19, 20].

3 HMIFS Model Description

In the following sections we develop more elaborate fractal systems based on hidden Markov model. The state of hidden Markov chain is $S_n \in \{X; \omega_1, \omega_2, \cdots, \omega_N\}$, the observation value of observed chain is the displayed way and color value of affine transformation points $\omega_i(x)$, $x \in X (i = 1,2, \cdots, N)$. Let:

$$P = (p_{ij})_{N \times N}, 1 \le i, j \le N \tag{3}$$

Where P is the state transition probability matrix of hidden Markov chain, the element p_{ij} is the transition probability from state ω_i to state ω_j, where:

$$p_{ij} = P(S_n = \omega_j \mid S_{n-1} = \omega_i), 1 \le i, j \le N,$$
$$p_{ij} \ge 0, \sum_{j=1}^{N} p_{ij} = 1, i = 1,2, \cdots, N \tag{4}$$

Let:

$$Q = (q_{jk})_{N\times M}, 1 \le j \le N, 1 \le k \le M \tag{5}$$

Where Q is the matrix which consists of several discrete probability value (Called observation probability) of the observation value $v_k \in \{O_m : m = 1,2,\cdots,M\}$ under the state ω_j, where:

$$q_{jk} = P(O_m = v_k \mid S_n = \omega_j), 1 \le j \le N, 1 \le k \le M,$$
$$q_{jk} \ge 0, \sum_{k=1}^{M} p_{jk} = 1, j = 1,2,\cdots,N \tag{6}$$

$\{X; \omega_1, \omega_2, \cdots, \omega_N, P, Q\}$ is called hidden Markov iterated function systems (HMIFS). HMIFS can generate more rich and colorful attractor image than MIFS by using Markov transition probability matrix and observation probability matrix.

When we use HMIFS system to construct fractal image, hidden Markov chain can be considered as the movements in fractal space. First , from a initial state begin, one of affine transformations is chose randomly according to a certain probability, and generate randomly the observations value $v_k (k = 1,2,\cdots,M)$ of affine transformation points $\omega_i(x)$, $x \in X (i = 1,2,\cdots,N)$,then, the model transform to the next state ,and continue to repeat this process. Finally, we can get two sequences by iterative calculation. One is the state sequence (can't be observed), the other is the sequence (can be observed) which consists of observation value of affine transformation points. There are two observation results, one is the displayed way of the sequence points, it decides the shape of images, the other is the color value of the sequence points, it decides the color of images.

3.1 HMIFS Fractal Modeling

Hidden Markov model is a dynamic model, it can transfer from one state to another state, it makes up a fractal image by some point sequences of affine transformation. These sequences are the observable object. We assume that every observable sequence point has the following one of 5 kinds of generating way. In different states, every different point's observation probability may be different. We can define the observation value of observable random process $\{O_m : m = 1,2,\cdots,M\}$ as follows:

v_1 denote grow, namely selecting affine transformation to do iterative calculation and draw the coordinates. The observation probability is q_{j1}.

v_2 denote end, namely the last point generated by affine transformation don't participate in the iterative calculation this time. The observation probability is q_{j2}.

v_3 denote pause, namely affine transformation only participate in the iterative calculation but don't draw the coordinates. The observation probability is q_{j3}.

v_4 denote reiteration, namely selecting affine transformation to do iterative calculation and draw the coordinates then return to the calculation. The observation probability is q_{j4}.

v_5 denote branch, namely affine transformation to do iterative calculation but don't draw the coordinates, draw other graphic instead of painting points. The observation probability is q_{j5}.

The shape observation probability matrix that consists of above observation probability $\{q_{j1}, q_{j2}, q_{j3}, q_{j4}, q_{j5}, j = 1,2,\cdots,N\}$ is $(q_{jk})_{N\times 5}, 1 \le j \le N, 1 \le k \le 5$, where:

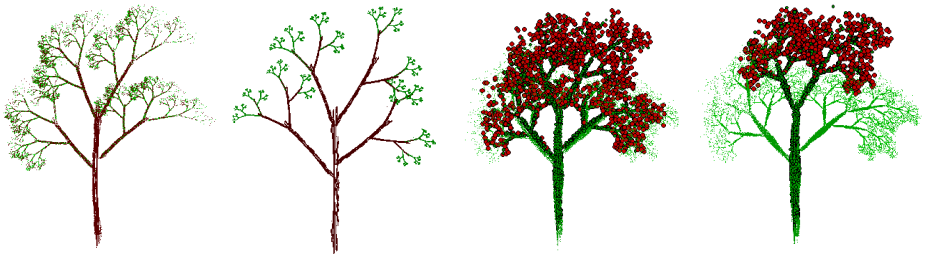
$$q_{jk} \ge 0, \sum_{k=1}^{5} p_{jk} = 1, j = 1,2,\cdots,N \tag{7}$$

In order to use HMIFS modeling fractal image, we choice freely the transition probability matrix of hidden markov chain and the observation probability matrix of observation chain, change easily the shape of HMIFS fractal image. Fig. 1 is the results of modeling to use different transition probability matrix and shape observation probability matrix, if we use red small round said the fruit of the trees, draw it to replace painting points, which simulate the fruit trees as fig. 1(c) and (d). Fig. 2 is the different deformation results of IFS square gaskets fractal.

$$P_1 = \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix} \quad P_2 = \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 0 & 1/6 & 2/6 \\ 1/6 & 1/6 & 2/6 & 2/6 & 0 & 0 \\ 1/6 & 1/6 & 0 & 0 & 2/6 & 2/6 \end{pmatrix}$$

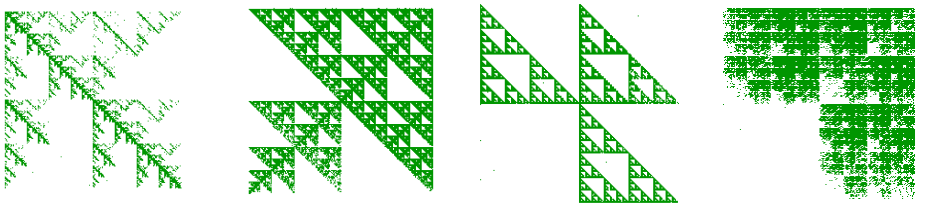
$$q_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \quad q_1 = \begin{pmatrix} 0.3 & 0.5 & 0 & 0.2 & 0 \\ 0.2 & 0.3 & 0.2 & 0.3 & 0 \\ 0 & 0.3 & 0.5 & 0.2 & 0 \\ 0.2 & 0.3 & 0.3 & 0.2 & 0 \\ 0 & 0.7 & 0.1 & 0.2 & 0 \\ 0 & 0.7 & 0 & 0.3 & 0 \end{pmatrix}$$

$$q_4 = \begin{pmatrix} 0.7 & 0 & 0 & 0.2 & 0.1 \\ 0.7 & 0 & 0 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.3 & 0.2 & 0 \\ 0 & 0.4 & 0 & 0.55 & 0.05 \\ 0.55 & 0.4 & 0 & 0 & 0.05 \\ 0.6 & 0.4 & 0 & 0 & 0 \end{pmatrix} \quad q_3 = \begin{pmatrix} 0.7 & 0 & 0 & 0.2 & 0.1 \\ 0.7 & 0 & 0 & 0.2 & 0.1 \\ 0.2 & 0.2 & 0.3 & 0.2 & 0.1 \\ 0 & 0.4 & 0 & 0.5 & 0.1 \\ 0.5 & 0.4 & 0 & 0 & 0.1 \\ 0.6 & 0.3 & 0 & 0 & 0.1 \end{pmatrix}$$



(a) matrix p_1, q_1 (b) matrix p_2, q_2 (c) matrix p_1, q_3 (d) matrix p_1, q_4

Fig. 1. HMIFS trees modeling



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

Fig. 2. HMIFS gaskets modeling

3.2 HMIFS Fractal Colour

In HMIFS system, if we define the observation value of sequence points of observable chain $\{O_m : m = 1,2,\cdots,M\}$ to be the color of the points then we can generate color fractal image.

v_1 denote Color value 1, the observation probability is q_{j1}.

v_2 denote Color value 2, the observation probability is q_{j2}.

……

v_M denote Color value M, the observation probability is q_{jM}.

The color observation probability matrix that consists of above observation probability $\{q_{j1}, q_{j2}, \cdots, q_{jM}\}$ is $(q_{jk})_{N\times M}, 1 \le j \le N, 1 \le k \le M$, where:

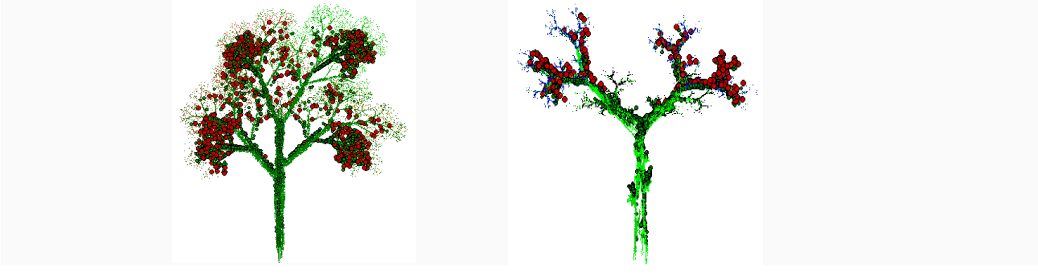
$$q_{jk} \ge 0, \quad \sum_{k=1}^{M} p_{jk} = 1, j = 1,2,\cdots,N \tag{8}$$

The color value of each affine transformation sequence point has one of M kinds of color (the size of the M according to need to set) in HMIFS images, it is controlled by the stochastic process which obey a probability distribution. We use a matrix $(c_{jk})_{N\times M}$ which has the same size with color observation probability matrix $(q_{jk})_{N\times M}$ to denote color distribution. The q_{jk} is the probability of color value c_{jk}. This kind of image system with color information has convenient colour function. We use transition probability matrix and color observation probability matrix to draw Fig. 3, we can find that HMIFS has rich and colorful color performance ability.

(a) (b) (c)

Fig. 3. HMIFS images colour

Fig. 4 shows the effect of modeling and colour which use comprehensive transition probability matrix, shape observation probability matrix and color observation probability matrix.



(a) (b)

Fig. 4. HMIFS images modeling and colour

4 Multi-level HMIFS Modeling

We use above the method to draw the global shape and color of IFS fractal images. In order to better show the images local details, we present a method of drawing with a multi-level local details as follow, it can model and colour separately for the local sub-images of image. First we use affine transformation of IFS system to do resolve, can get the resolving codes of the local sub-images. Then according to the method of section 3, we can generate more rich and colorful images.

Affine transformation resolving theorem: Let $\omega_1, \omega_2, \cdots, \omega_N$ be the IFS codes of fractal images. Let ω_i^{-1} be the inverse of ω_i. We define operator '$\circ$' to be $(\omega_i \circ \omega_j)(x) = \omega_i(\omega_j(x))$ .Then $\omega_i \circ \omega_j \circ \omega_i^{-1}, j = 1,2,\cdots,N$ is the IFS codes of the sub-images $\omega_i(f)$.

The matrix transformation formula for sub-images $\omega_i(f)$ as follows:

$$(\omega_i \circ \omega_j \circ \omega_i^{-1})(x) = A_i A_j A_i^{-1} x + (B_i + A_i B_j - A_i A_j A_i^{-1} B_i), (j = 1,2,\cdots,N). \quad (9)$$

According to the formula (9),we can resolve an IFS fractal image into several sub-images, continue the resolve the sub-images down into smaller sub-images, in this way, resolve step by step to the level of needs so far, the specific algorithm as follow:

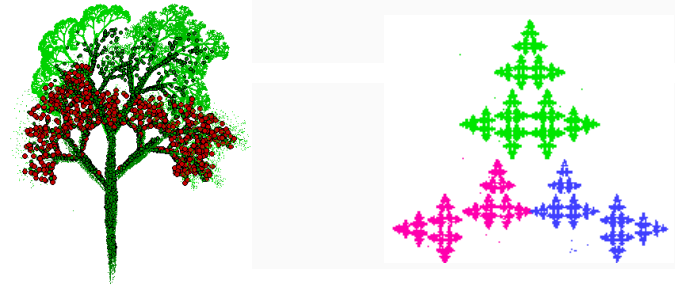
step1: we choose a few affine transformation ω_i $(i = j_1, j_2, \cdots, j_k, k \le N)$, use the formula (9), resolve the affine transformation ω_i , get the resolving codes of the sub-image $\omega_i(f)$.

step2: modeling the HMIFS $\{X; \omega_{ij}, j = 1,2,\cdots,N, P_i, Q_i\}$ of sub-images $\omega_i(f)$,

where $i = j_1, j_2, \cdots, j_k, k \le N$.

step3: output all the resolved HMIFS sub-images $\omega_i(f)$.

step4: output other unresolved HMIFS sub-images.



(a) (b)

Fig. 5. Multi-level HMIFS images modeling

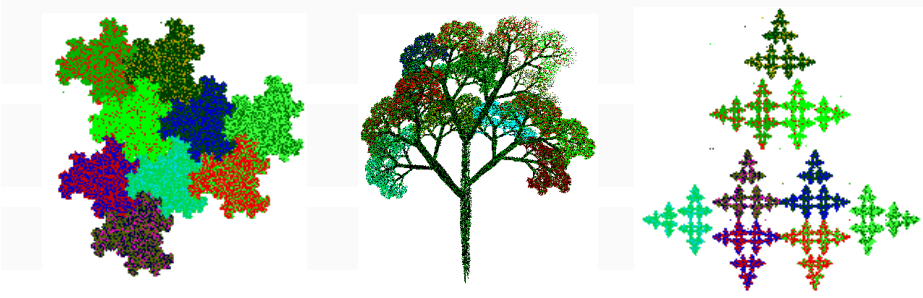


Fig. 6. Multi-level HMIFS images colour

We use multi-level HMIFS to draw Fig. 5. Fig. 6 which is the multi-level colouring's fractal images, showing that HMIFS is a general fractal images generation system.

5 Conclusion

An IFS fractal image is the congregation of affine transformation sequence points. The image characteristic is decided by the showing way of iteration points and color value. This article takes the hidden Markov model (HMM) into the IFS system, and establish the fractal image generation system HMIFS which is controlled by hidden Markov model. We applied probability distribution matrix to random modeling and colour for fractal images. It has the following advantages: First, the HMIFS use double random process to control attractor image generation, because of increasing control parameters, so it can create more complex and various images. Secondly, by choosing color distribution freely, it can draw the rich and colorful fractal images. Thirdly, using multi-level HMIFS model, it can not only control the shape and color of global structure, but can also control the shape and color of the local structure. It provides a new method for the natural scenery simulation, fractal art and graphic design, and virtual reality etc. We give the some experimental results of graphic example, but in practical application are not limited to this. It can be used in fractal image drawing of a more general type.

References

1. Darst, R., Palagallo, J., Price, T.: Fractal tilings in the plane. Mathematics Magazine 71(1), 12–23 (1998)
2. Adcock, B.M., Jones, K.C., Reiterect, C.A.: Iterated function system with symmetry in the hyperbolic plane. Computer & Graphics 24(3), 791–796 (2000)
3. Berthe, K., Yang, Y.: Optimization of fractal iterated function system with probability and fractal generation. Journal of University of Science and Technology 8(2), 152–156 (2001)
4. Andres, J., Fiser, J., Gabor, G.: Multivalued fractals. Chaos, Solitons and Fractals 24(3), 665–700 (2005)

5. Potgantwar, A.D., Bhirud, S.G.: Web enabled based face recognition using partitioned iterated function system. International Journal of Computer Applications 1(2), 30–35 (2010)
6. Nadia, M.G., Saidi, A.L.: On the security of image encoding based on fractal functions. International Journal on Computer Science and Engineering 3(1), 385–392 (2011)
7. Sprott, J.C.: Automatic generation of iterated function systems. Computer & Graphics 18(3), 417–425 (1994)
8. Karel, C., Simant, D.: Balancing order and chaos in image generation. Computer & Graphics 17(4), 465–486 (1993)
9. Martyn, T.: A new approach to morphing 2D affine IFS fractals. Computer & Graphics 28(2), 249–272 (2004)
10. Chang, H.T.: Arbitrary affine transformation and their composition effects for two-dimensional fractal sets. Image and Vision Computing 22(5), 1117–1127 (2004)
11. Dekking, F.M.: Recurrent Sets. Advances in Mathematic 44, 78–104 (1982)
12. Barnsely, M.F., Elton, J.H., Hard, D.P.: Recurrent iterated function systems. Constructive Approximation 5(2), 3–31 (1989)
13. John, W., Layman, T.: Linear Markov iterated function systems. Computer & Graphics 14(2), 343–353 (1990)
14. Hart, J.C.: Fractal image compression and recurrent iterated function systems. IEEE Computer Graphics and Application 5(7), 25–33 (1996)
15. Sherman, P., Hart, J.C.: Direct Manipulation of Recurrent Models. Computers & Graphics 27(5), 143–151 (2003)
16. Romberg, J.K., Choi, H., Baraniuk, R.: Bayesian tree structured image modeling using wavelet-domain hidden Markov model. In: Proceedings of SPIE, Denver, CO, pp. 31–44 (1999)
17. Xiao, L.L., Parizeau, M., Plamondon, R.: Training hidden Markov model with multiple observation-A combinatorial method. IEEE Transaction on Pattern Analysis and Machine Intelligence 22(4), 22–25 (2000)
18. Gough, J., Chothia, C.: Superfamily HMMs representing all proteins of known structure sequence searches alignments and genome assignments. Nucleic Acids Res. 30(1), 268–272 (2002)
19. Schlapbach, A., Bunke, H.: Using HMM based recognizers for writer identification and verification. Frontiers in Hand writing Recognition 17(1), 167–172 (2004)
20. Sun, J.X., Chen, Y.Z., Gun, D.B.: Multi-scale edge detection based on representation of edge characterized by hidden Markov model. Chinese Journal of Computers 26(4), 497–501 (2003)

High Performance Hybrid Streaming Scheme and VCR Support

Yong Wang* and Liangliang Hu

Department of Computer Science and Technology,
ChongQing University of Technology ChongQing, 400050, China
ywang@cqut.edu.cn

Abstract. In streaming media system, Static Streaming Scheme, like FB and GEBB, can reduce the usage of network bandwidth by using sharable multicast channel to deliver streaming, when most users access the same popular video. But Static Streaming Scheme has some problems: service delay, difficult to support interactive VCR and so on. To solve these problems, this paper proposes a new scheme HPHS and make some discussions to support interactive VCR. It merges the outstanding thoughts of HPSCT (revised from GEBB) and Patching scheme. The scheme aims at delivering popular video, allocates the number of channels adaptively according to request rate to minimize the network bandwidth consumption and satisfy the real-time playback requirement. Simulation results show that the scheme HPHS is high performance to decrease the usage of server bandwidth in Streaming media system.

Keywords: streaming scheme, VCR, GEBB, zero-delay.

1 Introduction

Streaming media system is typically implemented by client-server architecture, which assigns a dedicated unicast channel for each user. Masses of resources are required in the system. According to literature [1], the popularity of videos follows the Zipf distribution with the skew factor of 0.271, which means most of the demands (more than 80%) are for some (10% to 20%) very popular videos. Due to this fact, many broadcasting schemes are proposed, which can greatly reduce the requirements for network bandwidth. Pyramid Streaming Scheme [2] presented in 1995 was the first one. And from then on numerous optimized schemes have been put forward. These schemes can be divided to two classes: Static streaming scheme and Dynamic streaming scheme. The static streaming schemes divide a video to many segments and use multicast channel to broadcast them periodically, such as FB[3], GEBB[4], PHB[5], and so on. When users want to watch the video, they will have to wait for the beginning of the first segment. While they are watching the requested segment, the

* Project supported by Research Projects of Science and Technology Division from Chongqing Scientific and Technological Committee(No.CSTC2009AC2068).Science Research Starting Projects from Chongqing University of Technology(No.2008ZD24)

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 188–199, 2012.
© Springer-Verlag Berlin Heidelberg 2012

scheme can guarantee that the each segment is downloaded on time and the whole video can be played out continuously. Their main performance can be demonstrated with the relationship between network bandwidth assigned for broadcasting a video and the user waiting time. And these schemes have same problems: service delay and difficult to support interactive VCR. The Dynamic streaming schemes don't scheme the channel statically, and the assignment of multicast channel and unicast channel is dependent on the user arrival rate of the videos. Some typical schemes include patching [6] [7]. In these schemes, each new client will occupy a dedicated channel, which is not appropriate to support videos that are very popular, but it's zero-delay and easy to support interactive VCR.

In order to meet most users' requirement for minimal waiting time, Static Streaming Scheme usually assigns more bandwidth to hotter videos. But the hot degree of a video may change dynamically by many factors. So, some of revised schemes provide the solution to increase or decrease the number of channels seamlessly and dynamically in running, such as SFB[9], SSB[10] ,HPSCT [11].

To make the best of high performance of static streaming schemes and advantage of dynamic streaming schemes, this paper proposes a new hybrid streaming scheme HPHS, which absorbs the outstanding thoughts from HPSCT (revised from GEBB) and Patching scheme. The scheme use unicast channel to send the first segment of video, and use multicast channel to deliver other video segments. So, it can provide immediate service. It also can dynamically search the optimal number of channel assigned to the video by the newly updated arrival rate so as to minimize the bandwidth requirement.

The rest of this paper is structured as follows. In section II, the related video streaming schemes are introduced; then the new hybrid Scheme HPHS is presented in section III. In section V, we analysis and evaluate the performance of those schemes mentioned above. Conclusions are drawn in the last section.

2 Related Streaming Scheme

Streaming scheme must guarantee the users currently viewing a video will not experience any disruption. The static streaming scheme partitions a video into several segments, and periodically broadcast the each video segment on different multicast channels. No matter how the segments were downloaded from specifically channel by users, what should be emphasized is that the download time of each segment of the video must be ahead of or equal to its starting playing time

$$\text{DOWNLOAD\_TIME}(\varepsilon) \leq \text{PLAYING\_TIME}(\varepsilon) \quad (1)$$

Where ε stands for any part of video.theoretically.

2.1 GEBB Scheme

Consider a video of duration D, GEBB [4] scheme divide it into n segments $S_1,\cdots,S_n$; each segment S_i is repeatedly broadcasted in an exclusive channel with bandwidth $C = B/n$, where B is the total bandwidth for the video.

GEBB use the following equalities submitted to the formula (1) to compute the length of each segments. Given a user waiting time w (server delay) and the playback rate b.

$$S_i = \begin{cases} w*C & i = 1 \\ (w + \sum_{j=1}^{j=i-1} S_j /b)*C & 1 < i \le n \end{cases} \tag{2}$$

Given the D, B and n, the user waiting time w can be computed by the formula (3).

$$w = \frac{D}{(1 + B/(n*b))^n - 1} \tag{3}$$

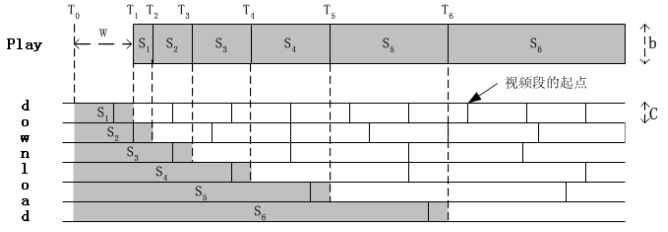


Fig. 1. GEBB scheme

Figure1 illustrates the process of downloading video segments on time and playing them at the user side, where the const service delay is w, the video is divided to 6 segments.

The streaming scheme of GEBB have special characteristics, which are the key points of achieving high performance: 1) Using lower channel bandwidth than video's playback rate; 2) Given a fixed total bandwidth, schemes have more channels with even bandwidth perform better; 3) Using greedy method to download data. Following the restriction of formula (1), GEBB gets the minimized length of the first video segment by maximizing the length of other video segments in other channels, so it shortens user-waiting time.

2.2 HPSCT Scheme

However, the length of video segments in GEBB has no corresponding relationship among using different number of channels (or total network bandwidth) to deliver a video. The character is important for changing the number of channels seamlessly in running. For high performance and realizing seamless channel transition with the waving of request arrive rate, changes, HPSCT[11] scheme is proposed.

2.2.1 General Description for HPSCT

This scheme aims at the broadcasting of a single video. The playback rate of a video V is b and the video is broadcasted using at least α ($\alpha \ge 1$) channels which bandwidth is b. When the video V has N ($N \ge \alpha$) channels with bandwidth is b, the partition method and streaming scheme are demonstrated as following:

Divide each channel which bandwidth is b into λ sub-channels equally. All sub-channels is show as $C_1, C_2, \cdots, C_{\lambda N}$; partition the video into λN video segments with different length $S_1, S_2, \cdots, S_{\lambda N}$ respectively. Any S_i is broadcasted on sub-channels C_i repeatedly. Depending on parameter α, the length of S_i (but S_1) follows rules as formula (4).

$$S_i = \begin{cases} (1 + 1/\lambda)\, S_{i-1} & 2 \le i \le \lambda\alpha \\ \sum_{j=1}^{j=\lambda\alpha} S_j / \lambda & \lambda\alpha + 1 \le i \le \lambda\alpha + \lambda \\ 2\, S_{i-\lambda} & \lambda(\alpha+1) < i \le \lambda N \end{cases} \qquad D = \sum_{i=1}^{\lambda\alpha} S_j \tag{4}$$

Table.1 shows the each video segment length and total video length (the length unit is the length of S_1) when $\alpha = 2$, $\lambda = 3$ and using different numbers of channels. The maximum possible length of S_i is calculated according to the GEBB scheme and the restriction of formula (1). Meanwhile the real length of S_i used in HPSCT is calculated by formula (4). And D is the sum of length of the video segments from the first sub-channel to the current sub-channel. From the masked data, the integral times relationship of video segment length is clear when using different number of channels to deliver the video.

Table 1. The length of video segment with $\alpha = 2$ and $\lambda = 3$

Channel	Sub-channel	Maximum possible S_i / S_1	Real length S_i / S_1	Total length D / S_1
1	C1	1.000	1.000	1.000
	C2	1.333	1.333	2.333
	C3	1.778	1.778	4.111
2	C4	2.370	2.370	6.481
	C5	3.160	3.160	9.642
	C6	4.214	4.214	13.856
3	C7	5.619	4.619	18.475
	C8	7.158	4.619	23.093
	C9	8.698	4.619	27.712
4	C10	10.237	9.237	36.949
	C11	13.316	9.237	46.187
	C12	16.396	9.237	55.424
5	C13	19.475	18.475	73.898
	C14	25.633	18.475	92.373
	C15	31.791	18.475	110.848
6	C16	37.949	36.949	147.797
	C17	50.266	36.949	184.746
	C18	62.582	36.949	221.696

The process of user downloading and playing each video segment is same as GEBB.

2.2.2 The Process of Seamless Channel Transition of HPSCT

The number of the segments and the number of channels need to be changed for lowest bandwidth consumption due to the waving of arrive rate in running. With the integral times length relationship of segments on the contiguous channel shown in the last column of table 1, the $\alpha+1$ channel and late channels are regard as some dedicated channels after transition, and the segments and delivering status in these channels don't need any change during the channel transition period. The only change is done in the first α channels and the increased or decreased channel.

The additional α channels are required to form new first α channels after the process of transition. When increasing one channel which bandwidth is b, the new channel will be placed after additional α channel, and each of old latter channel x (beginning from old α channel) is regarded as new channel $x+1$ after transition. Accordingly the segments in old first α channels are redistributed to new additional α channel and new increased channel synchronously. So, the user waiting time is shortened. If decreasing one channel, the segments in first α old channels and decreased the old channel (the $\alpha+1$ channel) will be distributed to additional α channels synchronously. Accordingly, each of old latter channe x (beginning from old $\alpha+1$ channel) is regarded as new channel $x-1$ after transition.[11]

2.3 Patching Scheme

Patching scheme allows a new user to receive its future playback data by joining an existing ongoing multicast channel, with the server using unicast channel to transmit missing data. So, all requests can be served immediately, the users can experience no service delay [6].

Under Patching scheme, the first request initiates a regular channel (sharable multicast channel), which will deliver an entire video repeatedly. For a new request for the same video, the server initiates a patching steam (unicast channel) which delivers only the missing part of the requested video. Without waiting for the next regular channel, the user downloads the data from the patching stream and immediately displays the data. Concurrently, the user downloads and caches the later portion of the video from the regular channel. Once the user starts to consume the data from the regular channel, the patching channel will be released.

Since the length of a patch stream increases as the age of the latest regular channel increases. It may be more efficient to start a new regular channel rather than to continue patching to the existing latest regular channel when the regular channel reaches a certain age [7]. Based on the idea, a more efficient scheme known as controlled multicast scheme was proposed [8]. In the scheme, a request arriving beyond T time units is served by initiating a new complete transmission for the video. When the user arrival for a video follows a Poisson process with the mean arrive rate λ and the length of the video is D in time, the optimal threshold is chosen to be $(\sqrt{2\lambda D+1}-1)/\lambda$. The required server bandwidth in units of the video play rate is $\sqrt{2\lambda D+1}-1$. The required server bandwidth grows with the square root of the user request rate[12].

2.4 Confluent Scheme of FB and Patching

A Workload Adaptive Scheme for Zero-Delay [12] is proposed to solve the user waiting time of fast broadcasting (FB) Scheme. It absorbs the outstanding thoughts from FB Scheme[3] and Patching scheme. To improve the performance, the scheme allocates adaptively the number of sharable multicast channels to deliver video segments periodically like FB Scheme. Using patching stream to deliver missing video data to user and provide zero-delay service. Both patching stream and multicast stream will consume network bandwidth. For optical total consumed bandwidth, the scheme revises the method of divided segment in FB to provide the capacity of seamlessly and dynamically performing the changing the number of channels. But the FB scheme is not the most bandwidth-effort static streaming scheme. We will show it in the Simulation results.

3 The New HPHS Scheme

3.1 Descript of the HPHS Scheme

Consider a popular video V (It also presents the length of video) with consumption rate b. The video is divided into k segments (the lengths and name as $S_0, S_1, \cdots, S_{k-1}$). Each segment S_i ($i > 0$ and $i < K$) is periodically broadcasted on its own sub-channel C_i with special bandwidth (normally is $1/\lambda$ the required playback rate, $\lambda > 1$). The $k-1$ sub-channels are called broadcast sub-channel. The S_0 is delivered via a unicast channel with bandwidth b (same as the required playback rate), like the patching stream in Patching scheme. Since all requests can be served immediately, the zero-delay service can be achieved.

For a new request to the video, the user downloads the S_0 from the patching channel (unicast channel) and immediately plays the video segment. Concurrently, the user downloads and caches all the other segments ($S_0, S_1, \cdots, S_{k-1}$) of the video from broadcast sub-channels. Once the customer consumes the S0, the patching channel will be released (or patching channel can serve other new coming user).In order to ensure the continuous playback, the customers should receive each video segments from the related broadcast channel right on time or before we need to consume them. So, segment S_1 should been downloaded completely while the playing process of S_0 has been finished, and it's same for other segments. To satisfy the requirement, we need exactly assign the length of ($S_0, S_1, \cdots, S_{k-1}$) and the bandwidth of each related broadcast channels.

Based on the analysis above, the rules on the server side are given to broadcast a popular video V:

1) According to the predicted popularity of the V, we assign N multicast channels with bandwidth b (same as the playback rate of video V) firstly. The number of channels can be increased and deceased dynamically in running according to the changing of user access rate. (It's same as HPSCT), but the least number of channels to deliver V is α ($\alpha \le N$, and it's const in running).

2) Divide each channel into λ sub-channels. So, there are sub-channel $C_1, C_2, \cdots, C_{\lambda N}$. Partition the video into $\lambda N+1$ video segments with length $S_0, S_1, \cdots, S_N$ respectively. Any of segment $S_i (i>0)$ will be transmitted on sub-channels C_i repeatedly. Depending on the least number of channels α, the length of S_i follows rules as formula (5).

$$S_i = \begin{cases} (S_0/b) * (b/\lambda) = S_0/\lambda & i = 1 \\ ((S_0/b) + \sum_{j=1}^{j=i-1} S_j/b) * b/\lambda = \sum_{j=0}^{j=i-1} S_j/\lambda & 2 \le i \le \lambda\alpha \\ \sum_{j=1}^{j=\lambda\alpha} S_j/b * b/\lambda = \sum_{j=0}^{j=\lambda\alpha} S_j/\lambda & \lambda\alpha+2 \le i \le \lambda\alpha+\lambda \\ 2\, S_{i-\lambda} & \lambda\alpha+1+\lambda \le i \le \lambda N \end{cases} \tag{5}$$

$$V = \sum_{i=0}^{\lambda N} S_i$$

Table.2 shows the length of each video segment and total video length (the length unit is the length of S_0) when $\alpha = 3, \lambda = 3$ with different numbers of channels. The maximum possible length of S_i is calculated according to the restriction of formula (1). Meanwhile the real length of S_i used in scheme of HPHS is calculated by formula (5), which back color is grey. And V is the sum of length of the video segments from the first channel to the current channel. From the masked data, the integral times relationship of video segment length is clear.

Table 2. The length of video segment with $\alpha = 3$ and $\lambda = 3$

Channel	Sub-channel unicast channel	Maximum possible S_i / S_1	Real length S_i / S_1	Total length V / S_1
0	**channel**	**1**	**1**	
1	C1	0.33	0.33	
	C2	0.44	0.44	
	C3	0.59	0.59	
2	C4	0.79	0.79	
	C5	1.05	1.05	
	C6	1.40	1.40	
3	C7	1.87	1.87	
	C8	2.50	2.50	
	C9	3.33	3.33	13.32
4	C10	4.44	4.44	
	C11	5.92	4.44	
	C12	7.40	4.44	26.64
5	C13	8.88	8.88	
	C14	11.84	8.88	
	C15	14.80	8.88	51.50
6	C16	17.76	17.76	
	C17	23.68	17.76	
	C18	29.60	17.76	102.33

3) For each new request to the video, the server starts a unicast patching stream to deliver the S_0. The user can play so immediately. Concurrently, the user downloads and caches all the other segments ($S_1,\cdots,S_{k-1}$) of the video from each multicast sub-channels. Once the user consumes the S_0, the patching channel will be released (the patching channel can serve other new coming user).So, the patching stream for one user stand the unicast channel shortly. When the users arrival for a video is mean, one continuous unicast channel can server many users.

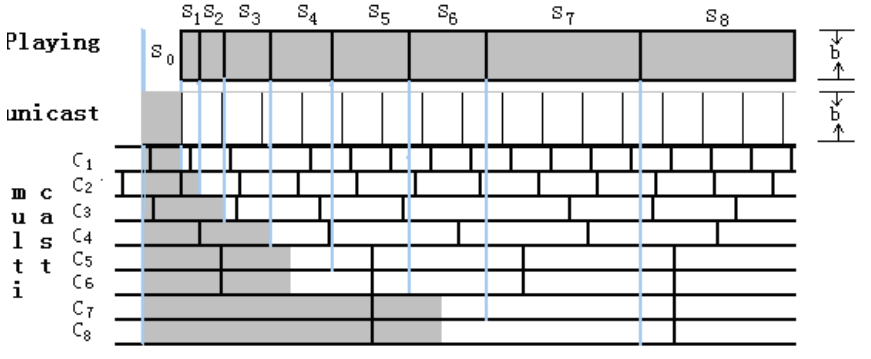


Fig. 2. HPHS scheme with $\lambda = 2, \alpha = 2$ and $N = 4$

4) The process of seamless channel transition in running
With the integral times length relationship of segments on the contiguous last multicast channels, the process of seamless channel transition is same as the description in 2.2(b).

3.2 Determination of Optimal Number of Multicast Channels for Different User Arrive Rate

In the scheme, the longer of the first segment, the longer of patching stream for each user and the more bandwidth consumed by patching streams(unicast stream) while the less multicast bandwidth is needed. The goal of scheme is to save total bandwidth consumption. So, how to assign the number of multicast channels and compute the length of the first segment is important.

Let b be the playback rate, a video's length is V, using N multicast channels which bandwidth is b. For a given const factor λ, The video be divided $\lambda N+1$ segments (the length is $S_0 \cdots S_{\lambda N+1}$). Let α as the least number of multicast channels to deliver the video.

1) For the given λ, α, N, the length of each segment follows rules as formula (5). We can conclude a function $S_0 / V = f(N)$.

2) The bandwidth consumption computation

Let x be the number of total users and B_p is the average bandwidth consumption by patching stream (bandwidth is b) during the period of playing the Video one time.

The patching period for each user is S_0 / b. Let B_m be the bandwidth consumption by multicast channel, and B_t be the total consumption. There are formulas as follows:

$$\begin{aligned} B_t &= B_p + B_m \\ B_p &= x * b * (S_0 / b) / (V / b) = x * b * S_0 / V \\ B_m &= N * b, \text{ So, } \quad B_t = x * S_0 + N * b = (x * f(N) + N) * b. \end{aligned} \tag{6}$$

The obtained average value of B_p from the equation (4) is a real number. But the bandwidth of patching stream is integer times of b in running. We have to round it to integer in simulation. For given const α, λ, the optimal N (To minimize B_t) is shown in figure 3 with different number of total users x. Like the factor n in GEBB, the const λ decides the bandwidth of sub-channels, and it can be big as possible (only restricted with how short the segment can be). Bigger λ can bring more bandwidth saving.

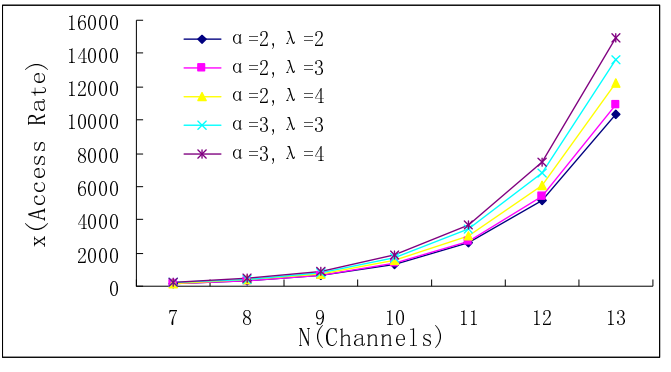


Fig. 3. Optimal N with different const factor λ, λ

3.3 Supports for VCR Interactivity

Users can initiate interactions after playing the video. There are some discussions about How to support these interactions.

A. Pause/Resume/Stop
These interactions are very simple to support. During Pause or Stop, the video can be downloaded and cached continuously. The required video segments can be found in local buffer when resume. So, there is no extra cost.

B. Fast Forward or Fast Reverse
Fast Reversed interaction can be resulted with local buffer. But to provide Fast Forward function, more unicast channel or special video coding is needed.

C. Jump Backward
If the video data after the destination position of Jump Backward have been cached, user can play it from buffer. Otherwise, it will be treated as Jump Forward interaction.

D. Jump Forward

Implementing unrestricted Jump Forward interactions will require additional bandwidth unless the video to be completely downloaded. A normal solution would be to start a unicast stream to transmit required video data which can't be received on time from the multicast channels for each Jump Forward request.

Server handles all destination point of Jump Forward which can be severed by local buffer, and these users being served by unicast stream. For some factors, we can group the users by requested access point of video, and start different multicast steaming to server different group. The main idea is that each multicast steaming is not a single channel with bandwidth b, but using HPHS scheme to broadcast the part video needed by a group user with many sub-channels which bandwidth is less than video playing rate b. The number of channels (total bandwidth) used to server a group can be computed by the description in 3.2. It will require much less additional bandwidth to meet the need of Jump Forward interactions.

However, it's complex that how to get the optimal group. It can be solved by an experiential function in real system.

4 Performance Comparison and Analysis

In our simulation, a CBR video of length of 90 minters is used. Let x stands for the number of total users arrived during the whole period of the video playing one times. We also suppose that the requests for the video arrive meanly. Fig.3 shows the optimal number of consumed multicast channels versus average user request rate x. As the user request rate increases, the number of segments also was increased to minimizing the total bandwidth consumption. The scheme can adjust N dynamically and seamlessly to adapt the changing of request rate.

In real system, inter-arrival of users is not mean, and the VCR interactions happen is randomly. They will bring the waving of the consumed bandwidth. We just show the average consumed bandwidth and don't count the bandwidth consumed by VCR interactions in simulation result. And the Fig.4 compares the required total bandwidth in units of the video play rate b with the Patching scheme [6] and the Confluent

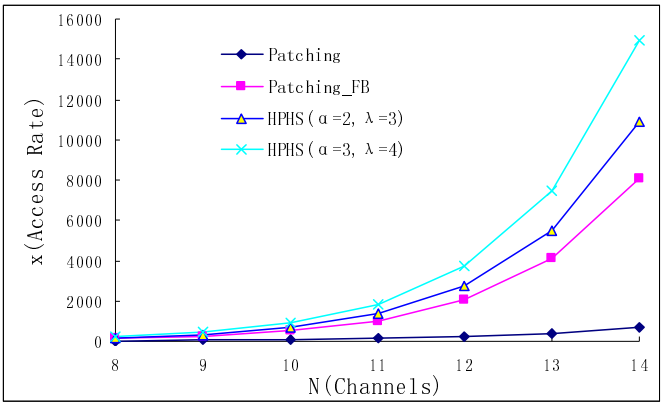


Fig. 4. Performance comparison

Scheme of FB and Patching(shown as Patching_FB) [12]. The proposed scheme has significantly lower bandwidth requirement than Patching scheme and Patching_FB[12] especially when the user request rate is high. In other words, HPHS can serve more users using the same total bandwidth. And the HPHS can get better performance with narrower sub-channels.

5 Conclusions

In streaming media system, using streaming schemes could save a large amount of network bandwidth. Allocating more bandwidth to deliver hotter videos could shorten the patching stream for each user and improve the server capacity of system. Therefore, we need a scheme, which is capable of dynamically and seamlessly changing the number of channels allocated to a video. To meets the demand, and solves the problem of start-delay in static stream schemes, a high performance hybrid streaming scheme is proposed with some discussions to provides the capacity of interactive VCR. The scheme combines the advantages of HPSCT scheme and patching scheme and can seamlessly perform the transition of changing the number of segments while guaranteeing clients currently viewing this video will not experience any disruption. Simulation results indicate that the scheme requires lower bandwidth than kindred schemes.

References

1. Dan, A., Sitaram, D., Shahabuddin, P.: Dynamic Batching Policies for an On-Demand Video Server. ACM Multimedia Systems (4), 112–121 (1996)
2. Viswanathan, S., Imielinski, T.: Pyramid Broadcasting for Video on Demand Service. In: Proceedings of MMCN 1995, San Jose, California, vol. 2417, pp. 66–77 (1995)
3. Juhn, L., Tseng, L.: Fast Broadcasting for Hot Video Access. In: Proceedings of Fourth International Workshop on Real-Time Computing Systems and Applications, pp. 237–243 (October 1997)
4. Hu, A., Nikolaidis, I., Beek, P.: On the Design of Efficient Video-on-Demand Broadcast Schedules. In: Proceedings of 7th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MAS-COTS 1999), pp. 262–269 (1999)
5. Paris, J.-F., Carter, S.W., Long, D.D.E.: A Low Bandwidth Broadcasting Protocol for Video on Demand. In: Proceedings of IC3N 1998, pp. 690–697 (1998)
6. Hua, K.A., Cai, Y., Sheu, S.: Patching: A multicast technique for true video-on-demand services. In: Proc. ACM Multimedia 1998, pp. 191–200 (1998)
7. Guan, D.L., Yu, S.Y.: A Two-Level Patching Scheme for Video-on-Demand Delivery. IEEE Trans. Broadcasting 50, 11–15 (2004)
8. Gao, L., Towsley, D.: Supplying instantaneous video-on-demand services using controlled multicast. In: Proc. IEEE Int. Conf. Multimedia Computing and Systems, vol. 2, pp. 117–121 (1999)
9. Tseng, Y., Hsieh, C., Yang, M., Liao, W., Sheu, J.: Data broadcasting and seamless channel transition for highly-demanded videos. In: Proceedings of INFOCOM 2000, pp. 4E-2 (2000)

10. Tseng, Y.C., Chueh, Y.C., Sheu, J.P.: Seamless channel transition for the staircase video broadcasting scheme. IEEE/ACM Transactions on Networking (TON) 12, 559–571 (2004)
11. Wang, Y., Gu, J., Jiang, K.-Z., Li, Z.-C.: High-Performance Seamless Channel Transition Video-Broadcasting Scheme. In: First International Conference on the Digital Society (ICDS 2007), p. 24 (2007)
12. Liu, Y., Yu, S., Wang, X.: A workload adaptive scheme for zero-delay video-on-demand service. In: Proceedings of SPIE, vol. 5960 (2006)
13. Fei, Z., Ammar, M.H., Kamel, I., Mukherjee, S.: Providing Interactive Functions through Active Client-Buffer Management in Partitioned Video Multicast VoD Systems. In: Rizzo, L., Fdida, S. (eds.) NGC 1999. LNCS, vol. 1736, pp. 152–169. Springer, Heidelberg (1999)

A Modular Image Search Engine Based on Key Words and Color Features

Xianying Huang* and Weiwei Chen

College of Computer Science & Engineering, Chongqing University of Technology, Chongqing,400054, China
{hxy,guaiguai_1015}@cqut.edu.cn

Abstract. Owing the widespread use of digital image, methods of high efficiency of image retrieval from WWW are becoming urgent requirements to users. But the traditional search engines are mostly based on keywords. This paper presents a modular image search engine based on keywords and contents, which organically combines search engine technology of keywords and images' color feature. The system searches images from WWW by WEB robots, extracts their relevant contents and color features, and then stores them into a database. When a user gives a query, the system displays the results according to the user's search requirements. For the color features of an image, a quantified method based on the maximum pixels ratio of irregular connected regions is raised. Experiments show that the method improves the retrieval efficiency and can get an expected search result more accurately, so as to satisfy the customer's needs.

Keywords: Image Search Engine, key words, color feature, architecture, feature extraction algorithm, connected area, quantify.

1 Instruction

With the growth of network bandwidth, the popularity of mass storage media and the rise of multimedia applications, the number and the capacity of the image increase at an alarming rate. Images are more and more used in the expression of information and the load of content. However, these images are distributed chaotically around the world and the information contained in the image cannot be accessed and used effectively [1]. Therefore, it becomes an important issue on image resource sharing that how to make it possible for people to obtain their desired image information quickly and accurately.

With the increasing demand of image search, a variety of WEB-based image search engines have appeared, such as Scour, Google, Yahoo, etc [5]. However, most existing image search engines are based on keywords and only a few can use visual properties for second search. Searching just from a series of keywords is not ideal for an image search, and the reason is that the descriptive texts for images are largely marked manually and different people have different views about the same picture, so

* Project supported by Chongqing Municipal Education Commission research(No.KJ10082).

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 200–209, 2012.
© Springer-Verlag Berlin Heidelberg 2012

they give the different descriptions about the same image. Thus, there is no uniform standard for describing images. As a result, the search results cannot satisfy the customer's requirements well.

The retrieval method based on image features and the content doesn't retrieve on the subjective assumptions, but on the information of the image itself. So it is objective. However, this method does not meet people's habit of querying. Therefore, considering the expansion, stability and maneuverability of the search engine, we use a modular way to organize the structure of the search engine, and a new kind of search engine based on the combination of keywords, image features and content is introduced in this article. Besides, the extractive method of color feature is studied deeply. Through this method, more expected results are gained accurately and the efficiency is good.

2 The Architecture of Image Search Engine

The overall architecture of a search engine, based both on keywords and image content features, includes the following several components: the module of image acquisition, the module of image feature extraction, the module of image database management and the module of user query, shown in Figure 1. The functional description of each module is described as follows.

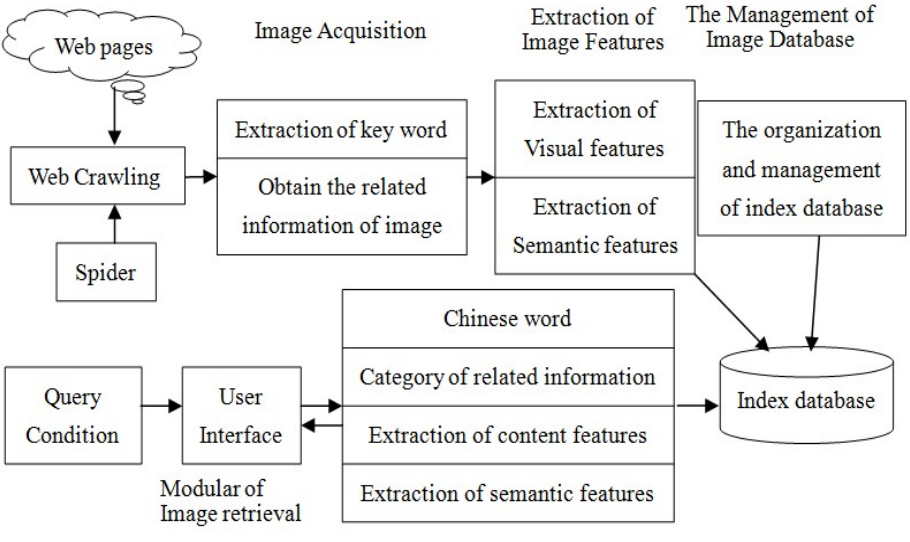


Fig. 1. Framework of image search engine

2.1 The Module of Image Acquisition

The function of this module is to obtain images from the network and to get the external information of images from web documents, such as size, description, link

text, etc. This is the beginning work of the search engine and is mainly finished by the network robot (spider). The spider starts from one URL or a set of initial URL list and reads a URL from the list constantly. Then it analyzes the Web page, and inserts the page's URL anchor into the list. Later, it chooses a new URL from the list, and sets it as the start. The spider repeats the process until there is no new URL in the list.

Spider use the IMG SRC and HERF tags in HTML to detect whether there are displayed image files. The IMG SRC tag means " The image file is shown here" and you can get the theme, keywords, size, link texts and other information of the images; the HERF means "Here is a link". The web page that locates the image and the URL of the image and the texts that surround the image are also important image information sources. Usually, examining the extension of a file can be used to determine whether the link is an image file or not. If the extension is GIF or JPG, then it is a displayed image. Spider then sends the image's URL link to the module of image feature extraction. After being extracted, the information is required to store into the database of image information, so as to facilitate the organization of index database.

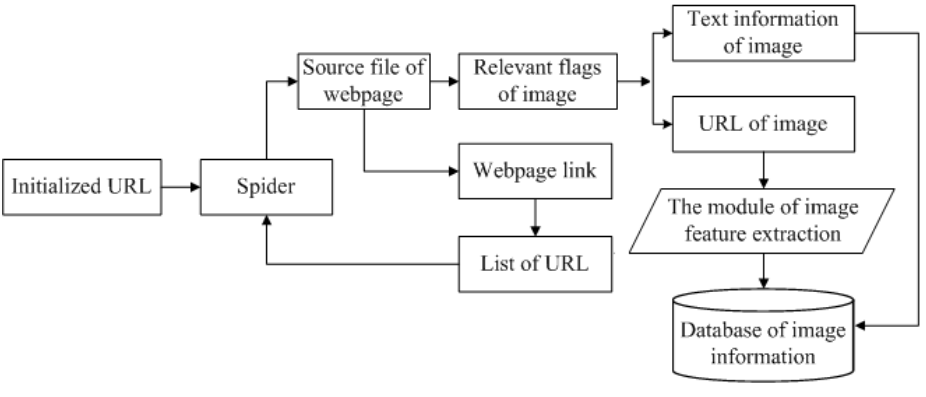


Fig. 2. The process of gathering images

2.2 The Module of Image Feature Extraction

The feature of the image is an original characteristic or the fundamental property, and it can distinguish one image from the others. Some characteristics correspond to the visual appearance of the image, such as color, shape, texture, etc., and can be used to calculator the similarity of the two images. This contributes a lot to the image search. The main function of this module is to extract image from its URL link, and extracts the image's feature information and then stores them into the database. Thus, it helps to organize to the index database later; shown in Figure 3.

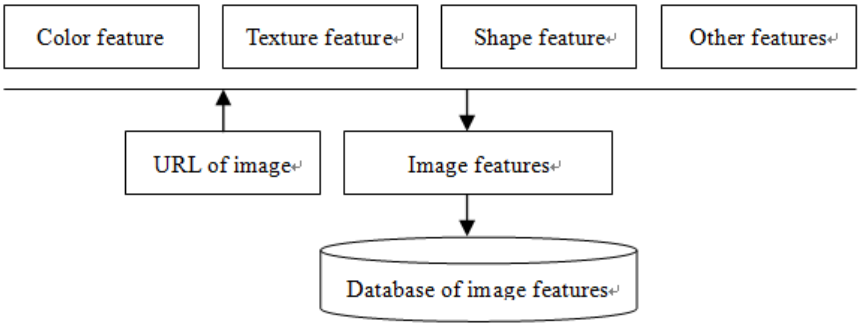


Fig. 3. Extraction of image content features

2.3 The Management Module of Image Database

The information stored in the image database should include the traditional index information, such as the keywords, the URL and the feature information. Following is an example data structure of an image database:

```
Struct picture_information //description of a picture
{
  String title;            //Subject of a picture
  String description;   //some detail information and Explanation
  Int width,hight;      //size of a picture
  Color_information color[9];//color detail information
}
Struct color_information   // color detail information structure
{
  Int color_value;      // amount of a color
  Boolean color_flag;//flag that use for identifying if a color is
prensent
  Real area_sum;// the amount sum of all area of this color
  Real area_max;//the maximum area of all area which has this
color in this picture
}
```

But most of the time, the database also stores the high-level features, such as semantic information. The common description methods of the image visual features are cumulative histogram, gray co-occurrence matrix and edge direction histogram. Therefore, when organizing the database, texts and related features of the images should be stored together. Besides, the convenience of the organization, management of image database, the future expansion and other factors should be also considered. For example, if the module of image-semantic retrieval is added, the structure of the image database should not be affected.

2.4 The Module of Image Retrieval

The main function of this module is to retrieve images that meet the user's requirements. Users can submit the query requests by the following ways: input the key words or input the image features (such as color, texture) or input an example image, etc. Users can select the single query mode, or the combinational one. In the combinational module, users can retrieval based on the first searching results. When searching by the way of example image and feature extraction methods, the system will call the feature extraction algorithm to extract the example image's features. Then it submits the feature vector to the search engine, and then compares the feature vector with the feature vector stored in the image database. At last, the system will present the result to the user in an inverted way according to the similar degree.

3 The Color Feature Extraction Algorithm

When users search by keywords and get a result, they can filter the result by the image features. They can also retrieve according to the given image features or example images. Since the color characteristics is the most obvious and intuitive features, we introduce the extraction algorithm of features with the example of color features below.

Color feature is a global feature, and it describes the surface property of the image or the regions of the image. It is a characteristic based on pixels, and each pixel of the image has its own contribution. The color feature is a visual feature which is the most widely used in image retrieval. This is mainly because that color is often related to the object or scene contained in the image and it has little relationship on the size, direction, perspective of image itself. So, it has a higher stability.

Before discussing the color feature, we should choose a proper color space to describe color. Here, we choose the CIELab color space. It is a descriptive method that it can turn light wavelength into brightness and hue, according to the characteristics of human vision, and is created by the International Lighting Association. It is also a well-proportioned color model and the difference between two colors corresponds two the Euclidean distance of the two color points in the color space. That is to say, the difference between to colors can be measured by space distance. In this color space, L represents the luminance of color; A represents the color scope ranging from green to red and B represents the color scope ranging from blue to yellow. Suppose(L_1,a_1,b_1) and$(L_2,a_2,b_2)$are color components of pixel c_1 and c_2 in CIELab color space, then the formula to calculate the distance between the two pixels is as follows:

$$E(c_1,c_2)=\sqrt{(L_1-L_2)^2+(a_1-a_2)^2+(b_1-b_2)^2}\,. \tag{1}$$

3.1 The Process of Quantifying Color

Color quantization means a process that abundant colors mapping to little colors. Thus, mistake exists inevitably. Suppose c_i is a three-dimensional vector in the color space, then C= {c_i, i=1, 2… N} represents the color set of image, during which N shows the number of color; in $\overline{C} = \{\overline{c_j}, j = 1,2,\ldots,K\}$, it represents the color set of color board, of which K represents the number of color in color board and K≤N. Then color quantization can be expressed as formula (2).

$$q : C \to \overline{C} . \tag{2}$$

When image quantification begins, we get the vector of each color in the image, and then classify it to the color which has the least difference in the color board. We can use formula (1) to calculator the distance between the given color vector and the color vector in the color board. The smaller the distance is, the closer the two colors are.

The process of extracting color is the process of quantifying image color. Now we give a general quantitative procedure firstly [2].

①Select a given number of points randomly in the image, supposing there are n points. Then take their colors $C_i(1 \le i \le n)$ as the value of each cluster center;

②Then, classify each pixel $c_j(1 \le j \le m)$, during which m means the number of pixels in the image. That is to calculator the color difference between each pixel and cluster center one by one, and then assign each pixel to the category that has the most similar color with it. The formula of calculating color difference between pixels can be expressed as follows.

$$D(c_j, C_i) = 1 - \exp[-\frac{E(c_j, C_i)}{\gamma}] \cdot \tag{3}$$

③Then, modify the color of cluster center. When one pixel is classified, the color of cluster center should be modified in time. The modified color of cluster center can be expressed by the average value of all pixels in the category. The formula can be written below.

$$C_i^{'} = \frac{1}{N} \sum_{1 \le j \le N} c_j \cdot \tag{4}$$

In the above formula, N means the number of pixels in the category.

④Repeat step ② and step ③ until convergence. That is, every pixel is classified and the color value of each category no longer changes.

⑤The split operation. Calculate the error for each category. If the error is lager than the pre-set threshold, then split the category into two categories.

⑥The merge operation. Calculate the distance between two category centers. If the distance of center of is smaller than the pre-set threshold, then merge these two categories into one.

⑦Then, calculate the quantitative results.

Suppose the color feature of image is expressed with set H and the quantitative color number of the image is expressed with n, then we can get the expression as follows.

$$H = \{(color_1, feature_1), (color_2, feature_2) \ldots (color_n, feature_n)\}$$

Replace the value of each image pixel with the corresponding component value of color feature, and then we will get the quantitative image.

3.2 The Algorithm of Quantifying Image

As there're many kinds of colors in the image, after being dealt with, the image can be divided into many regions, covered by different colors. Usually, the color distributes irregularly in the image. So, after being divided, the image turns to many irregular regions, with different colors. Besides, one color may cover discrete regions in the image. Thus, there're many regions of the same color.

Using the traditional method, we cannot calculator the area of the irregular region. Then, we can use the pixel ratio to express the result of image quantization. The pixel ratio of a region means the number of the pixels in the region divides the number of pixels in the image. The algorithm of quantifying image by pixels ratio is described as follows.

① Select n points randomly in the image, and takes each of them as the center point of a category and then get their color values as: C_1, $C_2 \ldots C_n$;

② For each center point C_i ($1 \le i \le n$), get the color value of its adjacent pixels in all directions, according to the region-growing algorithm. Then calculate the color difference between the pixel and the center. If the difference is the expected one, then spread in this direction further. Or, the point is the critical point in this direction.

③ Calculating in this way, the image will be divided into several irregular connected regions based on different colors and one color may also correspond to several regions in the image. Then, statistics the number of pixels of the maximum irregular region and calculate the number of pixels of the sum regions, covered by the same color. Suppose there're R blocks of region S, covered by color C_i, and region S_r($1 \le r \le R$) has N_r pixels. Then, we can calculator the number of pixels of the maximum region, covered by color C_i, using formula (5); we can also calculator the number of pixels of the total regions, covered by color C_i, using formula (6).

$$\max(C_i) = Max\{N_r, 1 \le r \le R\}. \tag{5}$$

$$sum(C_i) = \sum_{r=1}^{R} N_r \cdot \tag{6}$$

④ Calculate the pixel ratio of max (C_i) and sum (C_i), and express them by Pmax(C_i) and Psum(C_i) separately. The formulas are as follows, during which n means the number of color kinds.

$$Psum(C_i) = sum(C_i) \Big/ \sum_{i=1}^{n} sum(C_i) \cdot \tag{7}$$

$$P\max(C_i) = \max(C_i) \Big/ \sum_{i=1}^{n} sum(C_i) \cdot \tag{8}$$

⑤The results of image quantification are as follows.

$$H = \{(c_1, sum_{c1}, \max_{c1}), (c_2, sum_{c2}, \max_{c2}), ..., (c_n, sun_{cn}, \max_{cn})\} \tag{9}$$

When user searches, the results should be arrayed, and the standard of this quantization method is based on the pixel ratio. The larger the Psum(C_i) is, the more front the image is listed. When the Psum(C_i) of two image is almost the same, then the larger the Pmax(C_i) is , the more front the image is.

4 Experiments

According to the above-mentioned algorithm, a small image search engine based on keywords and color features is implemented. When the user searches with the keyword “Chongqing University of Technology”, the result is displayed as Figure 4, consistent with the theme.

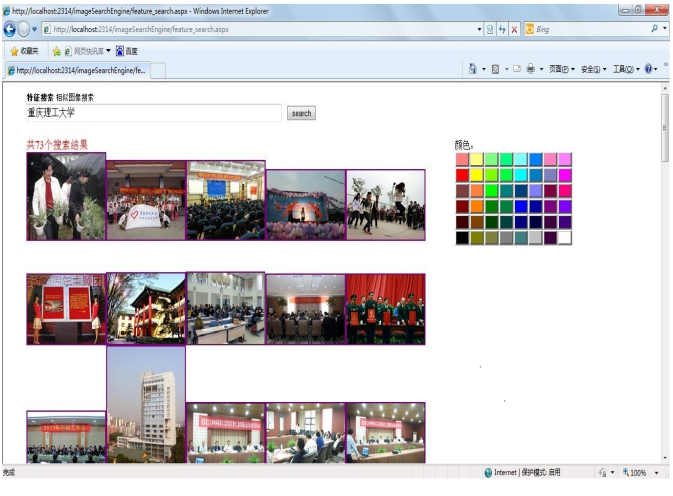


Fig. 4. Results searched with key word

When the user searches further and choose the red color, images with red as the main color are selected. For the traditional image search engines, users can also search for the second time. Figure 5 is a result, which is searched in “Baidu” search engine with the key word “Chongqing University of Technology” and the red color

for the second search. It's easy to find that, the result just display the images corresponding to the retrieval requirements, but not in a certain displaying order. The image search engine based on this algorithm can not only filter out images of color characteristics, but also sorts them according to color features, shown in Figure 6. We can see from the result that the larger the red area is, the more front its position is when it is arrayed. When the total area is roughly the same, the image that contains the largest separated red area has the priority to be arrayed. The result shows the correctness of the algorithm.

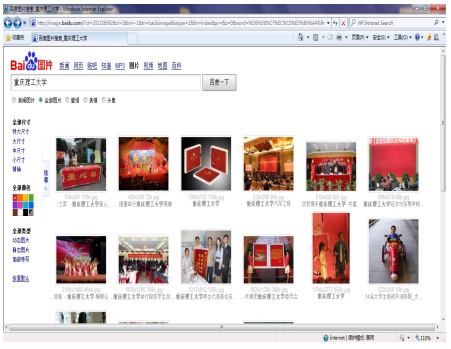


Fig. 5. Results of Baidu search engine

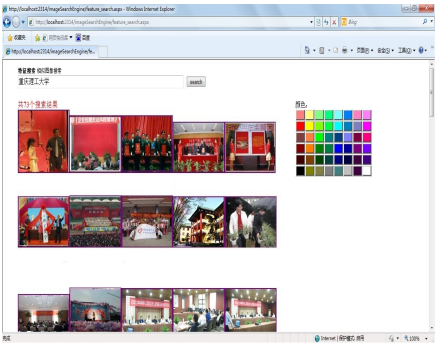


Fig.6. Results of personal search engine

5 Conclusions

The image search engine provides a great convenience for the users to retrieve the expected images from the big database on Internet. Traditional image search engine is based on the keywords and some on the image content also. But content-based search is based on high-level features or semantic (e.g., facial features, etc.), or on the underlying characteristics of image, such as color, texture, shape, spatial relations, etc. Thus, in this article, we design a modular framework of the search engine, and it can take various kinds of search methods or feature extraction ways into the system as separate modules, so as to increase the flexibility and scalability of the system. The color feature is the most obvious and intuitive features of the image, and this article introduces a method to quantify the color features in detail, using total area and the largest connected region with the same color. Experimental results show that this method is stable, and its retrieval performance significantly increases than the traditional histogram. So it can get the expected search results more accurately and can satisfy the customer's needs.

References

1. Wang, J., Liu, X., Liu, X.: Image Retrieval Method Based On Color Features and Relevant Feedback. Learned Journal of Shandong University of Technology (Natural Science) 23(3), 58–62 (2009)

2. Wang, W., Wang, Z.: Extraction Method Based On Content of Colorful Image Color Features. Computer Aided Design and Graphics Technology 13(6), 565–569 (2008)
3. Huang, Y., Guo, L.: An Image Retrieval Based on Object Region's Color and Texture Features. Journal of Nanjing University of Science and Technology 27(3), 286–289 (2009)
4. Fan, Y., Wang, R.S.: Color Image Segmentation Content Based Image Retrieval. Journal of Computer Research and Development 39(3), 376–381 (2008)
5. Bongiovanni, G.: Image Segmentation by A Multi-resolution Approach. Pattern Recognition 26(12), 18–27 (2009)
6. Malik, J., Belongie, F., Leugn, T.: Contour and Texture Analysis for Image Segmentation. Journal of Computer Vision 43(1), 7–27 (2011)
7. Eom, K.B.: Segmentation of Monochrome and Color Textures Using Moving Average Modeling Approach. Image and Vision Computing 17(3), 233–244 (2006)
8. Ning, Y.-D.: A Region Growing Algorithm Based on Color and Space Information. Computer Knowledge and Technology 5(12), 3196–3198 (2010)
9. Li, W., Huang, H.: The Application of Seeded Region Growing Technique in Color Image Segmentation. Micro Computer System 29(6), 1163–1167 (2010)
10. Chen, W.-B.: The performance comparison of some image similarity-matching methods. Journal of Computer Application 30(1), 98–100 (2011)
11. Zitova, B., Flusser, J.: Image registration methods: A survey. Image and Vision Computing 21(11), 977–1000 (2008)
12. Wang, X., Yang, H., Zheng, H., Wu, J.: Image Retrieval Algorithm Based On Block Color Histogram of Visual Weight. Automatic Technology, 1–4 (2008)
13. Fliekner, et al.: Query by image and video content: the QBIC system. IEEE Computer 28(9), 23–32 (2006)
14. Yang, Z., Peng, Y., Xiao, J.: Visual Vocabulary Optimization with Spatial Context for Image Annotation and Classification. In: Schoeffmann, K., Merialdo, B., Hauptmann, A.G., Ngo, C.-W., Andreopoulos, Y., Breiteneder, C. (eds.) MMM 2012. LNCS, vol. 7131, pp. 89–102. Springer, Heidelberg (2012)
15. Popescu, B., Iancu, A., Dan Burdescu, D., Brezovan, M., Ganea, E.: Evaluation of Image Segmentation Algorithms from the Perspective of Salient Region Detection. In: Blanc-Talon, J., Kleihorst, R., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2011. LNCS, vol. 6915, pp. 183–194. Springer, Heidelberg (2011)
16. Yang, Y., Zi, H., Shen, H., Zhou, X.: Mining multi-tag association for image tagging. LNCS, vol. 14(2), pp. 133–156 (2011)
17. Torjmen, M., Pinel-Sauvagnat, K., Boughanem, M.: Using Pseudo-Relevance Feedback to Improve Image Retrieval Results. In: Peters, C., Jijkoun, V., Mandl, T., Müller, H., Oard, D.W., Peñas, A., Petras, V., Santos, D. (eds.) CLEF 2007. LNCS, vol. 5152, pp. 665–673. Springer, Heidelberg (2008)

The Corroded Defect Rating System of Coating Material Based on Computer Vision

Gang Ji[1], Yehua Zhu[1], and Yongzhi Zhang[2]

[1] Chongqing University of Technology, Computer Science, City Chongqing, China, 400054
[2] No.59 Institute of the China Ordustry, City Chongqing, China, 400039
{jg,zhuyehua}@cqut.edu.cn, hszxzyz@126.com

Abstract. To solve the problem of artificial accurate detection of the coating material corroded defects and making accurate ratings automatically, this paper presents a method based on computer vision. To get the information of material corroded defects, the improved watershed segmentation method is used to eliminate over-segmentation. And for the processed image information, the parameter will be measured, and with the use of computer query technology, the information of the material corroded defects will be searched and contrasted to reach the function of machine rating.

Keywords: computer vision, Machine defect rating, image segmentation, coating corroded defects.

1 Introduction

A serious problem facing the world is the corrosion of material, the corrosion degree of material appearance is an important manifestation of the environmental adaptability of materials. Accurate detection and evaluation of the corrosion situation of material appearance is very important for the correct evaluation of the environmental adaptability of materials. In the trial of the natural environment, by the influence of environmental factors, the surface of coating material will be corroded, the corrosion characteristics including pitting, bubbling, peeling and so on will appear. Currently analysis of corrosion defect detection of coating material is mainly based on manual inspection and grade. Due to the differences of persons, detect errors often occur, and the missed check, error check may appear which will make it difficult to accurately detect and correct evaluation.

Computer vision which is used to replace the brain to process and explain is also a kind of input method using a variety of imaging system to replace the organ of vision. The ultimate goal of computer vision is to make the computer can observe and understand the world like human beings and have the independent ability to adapt to the environment. In the process of corrosion defect detection and rating of coating material, using computer vision technology, the information of corrosion defect of coating material will become digital and quantitative to facilitate the analysis and application, and also to improve the degree of standardization and accuracy.[1,2,3]

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 210–220, 2012.
© Springer-Verlag Berlin Heidelberg 2012

2 Computer Vision System and Defect Treatment

2.1 Structure of Computer Vision System

Due to the static measurement method of corroded defects detection of the coating material, to get the information of corrosion feature with computer vision system, the appearance image acquisition hardware system and related software system of corroded defects of the coating material can be set up with computer vision. The structure of the system is shown as in Fig 1.

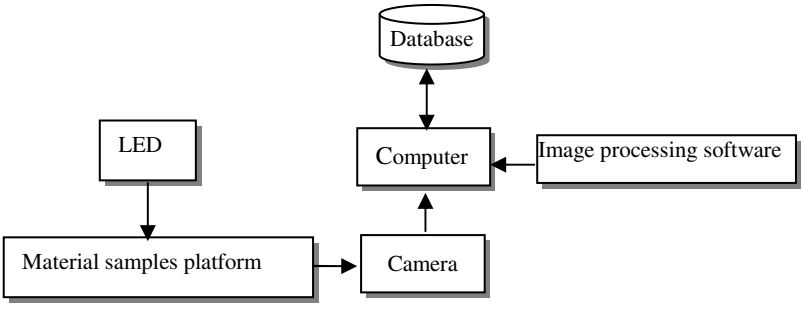


Fig. 1. Basic Structure Diagram of Computer Vision System

The material sample is placed on material sample platform and LED is used to achieve the best light effect, the image will be gathered with high resolution CCD camera and the information will be saved in image database, then the treatment and rating of coating sample will be received with image processing software. The image of corrosion sample of coating corrosion collected by the system is shown as in Fig 2[4,5,6].

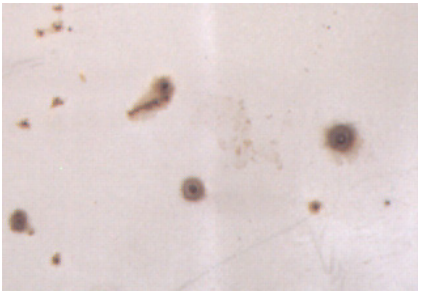


Fig. 2. Image Information of Corrosion Sample of Coating Material

2.2 The Process Mode of Corroded Defect Features of Coating Material

In the treatment of corrosion defect features of coating material, the collected image should be pretreated to improve the quality of images. The method such as filtering, passivation, contrast stretches, clipping can be used to get the information of features. The process of image character is based on the understanding of image, the interested target in image which is described in the computer should be divided and the information can be gotten.

The goal of the segmentation of corroded defects in coating material is to identify corrosion characteristic information of coating material. [7,8,9] Image segmentation is the key step in the image analysis which for many years get attention of people highly and has been widely studied in many fields. The method of segmenting includes the method based on threshold segmentation, the method based on edge detection, the method based on region and the method of watershed segmentation based on the morphology, but in the actual application the ideal effect on all the image information is hard to get and there is still no unified standard to judge one segmentation algorithm is good or not. For the corrosion characteristic information of coating material, segmentation method conforming to the image characteristics should be studied and designed based on characteristics of original image and with priori knowledge to improve the quality of image segmentation. In this paper the improved watershed segmentation method is applied to process image.

3 Process of Corrosion Characteristic Information of Coating Material Based on Watershed Segmentation

3.1 The Treatment of Watershed Transformation

Watershed transformation is a classical algorithm in topological field. The basic idea is that the image is regarded as geodetic topology landscapes, the brightness of every pixel of image represents the altitude of the pixel, each local minimum and its influence area is called reception basins, which is shown as in Fig 3.

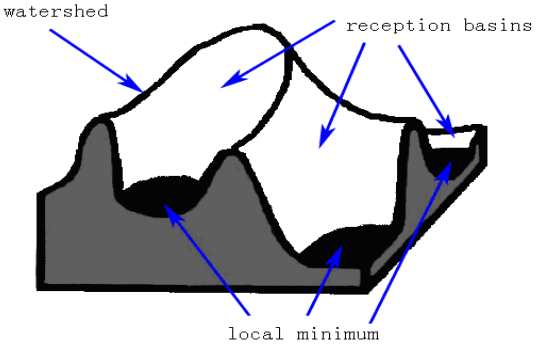


Fig. 3. Watershed Schematic

The goal of watershed segmentation algorithm is to extract information from digital image, in many of the watershed transform algorithm, Immersion simulation algorithm[34] offered by Vincent and Soille in 1991 is the most famous and the fastest solution to the calculation of the watershed transformation algorithm. The idea of the algorithm is that making tiny holes in the surface of each local minimum in the image then put the whole image model slowly dipped into the water. With the deepening of immersed, the influence of each local minimum expands out slowly. When the water level of two adjacent reception basins will rendezvous, a dam will be built to prevent the meet at the confluence of the two adjacent reception basins. At the end of the immersion, each area is flooded and completely surrounded by the dam, the collection of the dam comprise watershed of the image. The algorithm mainly includes sequence and flood.

(1) How to sort.
Firstly the whole image will be scanned to get probability density of all the gray level and the position of each pixel in the sorted array which can be gotten through calculating cumulative probability of gray distribution and the gray value of the pixel. All pixels will be saved in the sorted array with the standard that the smaller the gray value is the more ahead position of the pixel is.

(2) How to flood.

1) Pixels will be sorted in order of gray value from low to high and the pixels with same gray value will be regarded as the same gray level.
2) Deal with a gray level named hcur (the current layer), the pixel which's neighborhood has been marked will be added to a FIFO queue.
3) If the FIFO queue is not empty, the first element will pop up and become the current pixel being processed. Consecutive points of the current pixel which's height is variable hi will be processed in sequence. If consecutive point has been marked, then the current pixel will be refreshed according to the mark of the consecutive point; If consecutive point has not been marked, then add this consecutive point to the FIFO queue. Repeat this cyclic step until queue is empty.
4) Again scan the pixel of current gray level, check whether there are points which are still not marked. Because the points not marked means a new minimal district, so if there are points not marked, the sentinel value of current region add 1 and perform as the same step as 3) to mark all the pixels of this minimum area.
5) Return to step 2 and process the next gray level until all gray level is handled and closed.

In the above algorithms each pixel is scanned 5 times averagely, sorting twice and flooded three times, so the execution time is linear.

In fact the grads (∇f) which is a vector of image describes the gray variation. Definition for grads is:

$$\nabla f = \left(\frac{\delta f}{\delta x}, \frac{\delta f}{\delta y}\right)$$

Definition for grads module which determines the value of grads is:

$$|| \nabla f \ || = \sqrt{(\frac{\delta f}{\delta x})^2 + (\frac{\delta f}{\delta y})^2}$$

The grads of a grayscale describes the grayscale of the image. The smaller the change of gray level is, the smaller the grads is, so the local minimum value of reception basins can be related to the target with the smaller grads value. In the traditional watershed segmentation method the grads image is extracted by gradient operator, and then the grads image is decomposed by the watershed transformation. Affected by the impact of noise and the weak edge the interior gray of object varies in degree and there are regional minimum in the grads image respectively. In the principle of watershed transformation, each independent regional minimum of the input image produces different area, so the number of image region with the watershed transformation is same as the number of regional minimum in the input image. So the slight change of grads will produce a lot of meaningless small area after the watershed transformation, usually the correct contour information is submerged in complex edge information which is known as "the over-segmentation" and useful target feature information can not be accurately extracted[10,11,12,13,14]. Figure 4 shows the over-segmentation of rust feature.

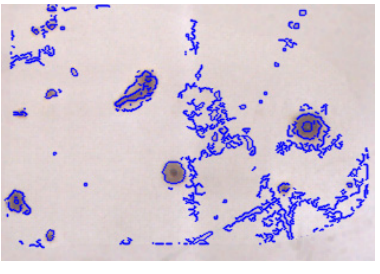


Fig. 4. Over-Segmentation of Watershed Transformation

3.2 The Improved Watershed Transformation

Watershed transformation is an effective segmentation method to extract the regional feature, the closed and continuous edge can be acquired, which can bring a lot convenience for analyzing the size and area of corrosion characteristics. Since traditional watershed transformation is very sensitive to the noise and weak edges that can easily lead to over-segmentation, the better results will be achieved only when the method directly act on those images that have no noise and symmetrical and stable grads. For the image of corrosion features of coating material the traditional watershed transformation must be improved to get better result.

There is two way to solve the problem of over-segmentation. The one is to use the watershed transformation firstly, and then merge the adjacent region with some rule. Another way is to do some corresponding preprocessing and extract tags of interested target before segmentation, mark those that have tags as the minima of grads image

and block the minima not marked to eliminate the noise and then use the watershed transformation to get the target that has been marked.

In order to reduce the influence of over-segmentation the following improved method is proposed which use the gray information to get the initial partition and then merge the part of over-segmentation by using the color information. The specific steps are shown as follows:

1) Use the formula 3.1 and 3.2 to extract the value of color (H) and brightness (I) from the RGB image.

$$H = \begin{cases} \theta & B \le G \\ 360\text{-}\theta & B > G \end{cases} \tag{3.1}$$

$$I = 0.299R + 0.587G + 0.114B \tag{3.2}$$

In formula 3.1 the θ is:

$$\theta = \arccos\left\{ \frac{\frac{1}{2}[(R-G)+(R+G)]}{[(R-G)^2+(R-G)(G-B)]^{\frac{1}{2}}} \right\} \tag{3.3}$$

2) For I the Canny operator is used to get grads image and correct the image with threshold, so make the reception basins response as far as possible to the interested target.

3) Split the modified grads image with the watershed detection algorithm proposed by Vincent.[15,16,17,18,19]

4) The image is divided into different small regions after the watershed transformation. Due to the impact of weak edges the adjacent regions need to be merged. During the merger, the regional difference determines whether two adjacent regions should be merged. The region difference is defined as the Euclidean distance of tonal average of region.

$$d(R_i, R_j) = \left| \mu_{R_i} - \mu_{R_j} \right| \tag{3. 4}$$

Ri, Rj represent i-th and j-th region, μ_{R_i}, μ_{R_j} is the corresponding tonal average of region.

In this method, the grads and color information is considered comprehensively. The processing result of rust and peeling of coating material with this improved method is shown in Fig 5 and Fig 6. In application because of the diversity of the corrosion characteristics of coating materials the threshold of grads correction and merging is changeable, so the human-computer interaction is used to aid in correcting the interested regions to obtain more accurate and satisfied results.

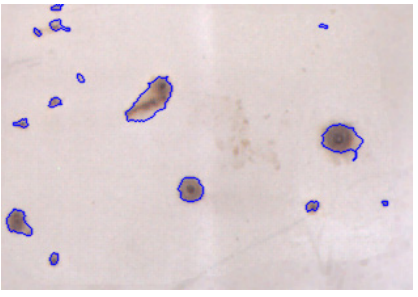


Fig. 5. Segmentation Result of Rust

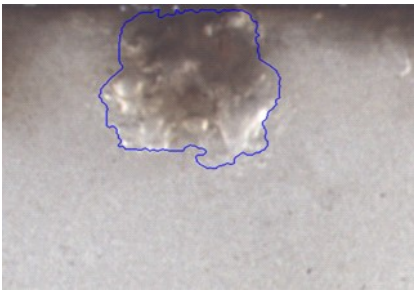


Fig. 6. Segmentation Result of Peeling

4 The Rating of Corroded Defect of Coating Material

4.1 Measurement of Corrosion Characteristic Parameter

The grading of corrosion defects of coating material is to calculate and rate the quantity and size of corrosion feature by getting the information of corrosion characteristics, so the feature parameters of corrosion must be calibrated and measured firstly. Because image gotten by digital camera is based on pixels the relationship between the size of pixels and the trim size must be built which is called calibration parameter.

The tool to get calibration parameter is a standard calibration board with precise photolithography rulers which is shown in Fig 7.

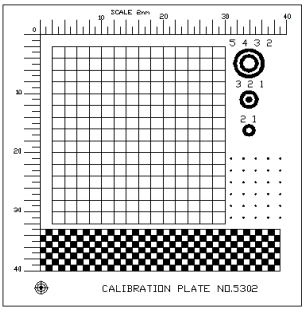


Fig. 7. Schematic of Calibration Board

Steps of the process are:

1) Collect image of the board to get information of the board.

2) Measure the size of unit distance recorded as L has how many pixels to get calibration parameter recorded as K. If the shooting is under the condition that the focus and object distance is fixed, the relationship between L and D is shown as follows:

$$L = K \times D + e \tag{4.1}$$

In formula 4.1 e is the system error.

The two group of measurement value of L and D will be gotten recorded as D1, L1 and D2, L2 and substitute the value for formula 4.1, so the K and e can be calculated.

$$K = \frac{L_2 - L_1}{M_2 - M_1} \tag{4.2}$$

$$e = L_1 - \frac{L_2 - L_1}{M_2 - M_1} \times M_1 \tag{4.3}$$

4.2 The Measurement of Corrosion Characteristic Parameter

(1) The measurement of the size of corrosion character

Usually the shape of the corrosion spots are irregular and the maximum diameter calculated by the minimum circle-cover of contour points of the corrosion features can be used to represent the size of corrosion features. The minimum circle-cover is defined as the circle with the smallest radius which contains all the points of a given set called as P.

(2) Count the number of the corrosion character

The single-pixel wide, connected and closed region can be received by the improved watershed transform. Through finding and counting up all the closed regions the number of the corrosion character will be gotten. The original contours can be found in order from left to right and from top to bottom.

(3)Calculate the area and area percentage of the corrosion character

Count the total number of pixels of all enclosed area of corrosion character which can be used as the area of corrosion character. Area percentage can be calculated by the following formula:

$$\frac{\sum_{i=1}^{n} A_i}{A} \times 100\% \tag{4.4}$$

In the formula, A is total area of the image, A_i is area of the i-th corrosion point, n is the count of corrosion points.

4.3 The Grading of Corroded Defects of Coating Material

After segmentation of the corrosion characteristic and the measurement of corrosion characteristic parameter, compare corrosion characteristic with the rating standard of coating corrosion with computer query technology and the result can be acquired. [20,21,22,23,24]

(1) The rating content of corroded defects of coating material
The corroded defect features of coating material can be measured by the number, size, area and so on, according to the standard GB / T 1766-2008 the key information about the includes:

1) Rating of rust
   The number of rust spots, maximum size of rust (mm), rust area (mm2) and rust percentage of total area can be the standard of rating. The rating standard of rust is shown in Table I.

Table 1. The Rust Rating Standard of Coating Material

standard GB / T 1766-2008		Image Rating
Grade	Rust size (maximum size)	
S0	no visible rust under 10 times magnification	0
S1	the rust is visible only under 10 times magnification	size of rust is <0.1mm
S2	the rust is visible just under normal vision	size of rust is 0.1mm ~ <0.2mm
S3	size of rust is <0.5mm	size of rust is 0.2mm ~ <0.5mm
S4	size of rust is 0.5mm~5mm	size of rust is 0.5 ~ 5mm
S5	size of rust is >5mm	size of rust is > 5mm

2) Rating of bubbling
   The number of bubbling, maximum size of bubbling (mm), bubbling area (mm2), and bubbling percentage of total area can be the standard of rating. The rating standard of bubbling is elided.
3) Rating of peeling
   The number of peeling, maximum size of peeling (mm), peeling area (mm2), and peeling percentage of total area can be the standard of rating. The rating standard of peeling is elided.

(2) Steps to rate

1) Establish the rating standard database of corroded defect of coating material.
2) Process and measure the corroded defect information of the coating material.
3) With the computer query technology the corroded defect information of the coating material is used as the motherboard (query keywords) to get the grade of corroded defect of the coating material by searching in the rating standard database.

5 Conclusion

The image processing of corroded characteristics of coating material is an important method to achieve the corroded characteristics of coating material, the segmentation and calculation of corroded points is the key for corroded characteristic detection and rating. With Visual C + + 6.0 development tool functions such as image segmentation, rating of corroded characteristic and so on has been completed which have been applied in practical application and can completely replace the manual rating. With the development of computer technology and image processing technology, the research of this system will be more deeply into the application areas of material corrosion and will become a new direction of corrosion science.

References

1. Ji, G., Fei, Z., Zhu, X.-C.: Application of Information Hiding Algorithm of Secondary Embedding Based on LSB in The Process of Corrosion Material Image. Journal of Chongqing Institute of Technology (Natural Science Edition) 24(7) (2010)
2. Ji, G., Zhang, J.-X.: The Inspection System of Material Decay Original Value Based on Image Query Technology. Journal of Chongqing Construction University, EI Search 27(4), 125–128 (2005)
3. Ji, G.: On-line detection system for corrosion features on the surface of materials based on computer vision technology. Computer Science 4B (2009)
4. Jung, C.R.: Combining wavelets and watersheds for robust multiscale image segmentation. Image and Vision Computing (25), 24–33 (2007)
5. Ji, G.: Inspection System for Original Value of Material Decay Base. Journal of Chongqing Construction University (4), 125–128 (2005)
6. Zhang, J.-X., Ji, G.: Inspection System for Original Value of Material Decay Based on Image Search Technology. Computer Engineering (16), 202–204 (2007)
7. Sonka, M., Hlavac, V., Boyle, R.: Image processing, Analysis and Machine vision. People's Post and Telecommunication Publishing House, Beijing (2003)
8. Qu, Y.-P.: Application of Image Recognition Technology in Evaluating The Color of Material Corrosion Image. Surface Technology 34(4), 71–72, 75 (2005)
9. Yan, X., Sun, X.-S.: Application of The Image Processing Technology in The Metal Plate Corrosion Test. Electronic Technology (3), 23–25 (2009)
10. Sahoo, P.K., Arorab, G.: A Thresholding Method Based on Two-dimensional Renyi's entropy. Pattern Recognition 37, 1149–1161 (2004)
11. Yan, F., Zhang, H., Ronald Kube, C.: A multistage adaptive thresholding method. Pattern Recognition Letters (26), 1183–1191 (2005)
12. Soille, P.: Morphological Image Analysis: Principles and Applications. Tsinghua University Press, Beijing (2008)
13. Cai, D.: Application of Watershed Transform in Image Segmentation Based on Mathematical Morphology. Journal of South-Central University for Nationalities (Natural Science Edition) (4) (2005)
14. Navon, E., Miller, O., Averbuch, M.: Color Image Segmentation Based on Adaptive Local Thresholds. Image and Vision Computing (23), 69–85 (2005)

15. Wang, X.-P., Hao, C.-Y., Fan, Y.-Y.: Watershed Segmentation Based on Morphological Scale-Space and Gradient Modification. Journal of Electronics and Information Technology 28(3), 485–489 (2006)
16. Yang, W.-L., Guo, L.: Image segmentation method based on watersheds and graph theory. Computer Engineering and Applications 43(7), 28–30 (2007)
17. Yu, Y., Qi, Z.: Image Semi-segmentation. Journal of Communication University of China Science and Technology (Natural Science Edition) 16(3), 26–31 (2008)
18. de Berg, M.: Computational Geometry:Algorithms and Applications. Tsinghua University Press, Beijing (2005)
19. Ji, G.: Research on the Processing and Appraising System for Material Appearance Erosion Characteristics. Journal of Chongqing Institute of Technology (Natural Science Edition) (1) (2007)
20. Ji, G., Fei, Z., Wang, P.: Research on Semantic Annotating of Corrosive Material Image base on Ontology. Journal of Chongqing Institute of Technology (4) (2011)
21. Zhang, Y.-J.: Image Processing and Analysis Techniques. Higher Education Press, Beijing (2008)
22. Ji, G.: DICOM Image Processing Technology in CT. Computer Applications and Software (Core Journal) 24(11) (2007)
23. Ji, G.: Research on the Processing and Appraising System for Material Appearance Erosion Characteristics. Journal of Chongqing Institute of Technology (Natural Science Edition) 21(1) (2007)
24. Zhou, X.-C.: Study of Image Segmentation Methods and Their Applications. Journal of Image and Graphics, 1–10 (2005)

Service Replacement Algorithm Using Complex Network

Yang Zhang[1] and Chuanyun Xu[2]

[1] College of Computer and Information Science, Chongqing Normal University, Chongqing 400047, China
[2] College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400050, China
495461428@qq.com

Abstract. Aimed at the problem that the results of services replacement will be affected by complex dependency relations in services systems, the dependency relationship between services in the service system is formally analyzed, and Service Replacement Algorithm using Complex Network (SRA-CN) is proposed. In order to discover and select the best new services, SRA-CN use the replaceable degree got from the dependency relationship to estimate the affect depth & affect breadth of dependencies between services to the replacement process and its results. At the end, two groups of experiments between SRA-CN and its correlation algorithm are performed to analyze the performance of SRA-CN.

Keywords: Web Services Replacement, Dependency Relationship, Replaceable Degree, Complex Network.

1 Introduction

Because the functions of single web service could not meet the requirements of complex application, several services with different functions should be composed to form a new powerful service. Traditional process of web services composition follow "Definition before Execution", which means that the static process of web services composition which meets the business requirements is pre-established by exports. At present, the researches about web services composition are mainly based on automated reasoning technology. For example, the standard business process modeling of web services called by users are done by XML Service Request Language planning framework to plan & operate services process alternately [12].

The network formed by many adaptive web services could be seen as a complex system which would emerge some structures and functions by the nonlinear interactions between individuals under the conditions of no central control, incomplete information and mere existence of local interactions. Complex network is the backbone of a complex system which would be taken as an interaction network between units or individuals. In traditional methods, the network structure is often described as a completely random graph that is E-R random model made by Erdos and Renyi [13]. But with further research on different complex networks in reality (Eg: Internet, World Wide Web, social networks, citation networks of scientific

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 221–231, 2012.
© Springer-Verlag Berlin Heidelberg 2012

papers, etc.), many phenomena cannot be described by E-R random model. In order to describe the transformation process from a conventional network to a random network, Watts and Strogatz propose a new "Small-World" Network Model [14,15] (WS Network Model), in which any linking in an adjacent node ring interconnection network would re-establish a connection with other nodes in accordance with a certain probability. The other major discovery is "Scale-Free" Network Model proposed by Barabasi and Albert [16], in which new nodes are constantly added to the network and select the network nodes with a large degree to establish a connection. But the network structures show inconsistencies in reality [17], because most nodes in the network (such as the Internet, World Wide Web, etc.) have only a few connection as well as very few nodes have a large number of connections and the node degree show some exponential characteristics. Hence we use Scale-Free Network to describe web services flows in the composition with a proper node degree. At present, the mathematical tools of complex networks contain graph theory, combinatorial optimization, statistics, probability and random processes. In this paper, we choose genetic algorithm as the model algorithm of scale-free network for its good optimal performance. Genetic Algorithm is a kind of nonnumeric vale calculation optimum method built on the basis of natural selection and group genetics, it simulates parents propagate manner of great majority nature creatures using nature selection mechanism of “Select the Superior and Eliminate the Inferior” mainly by crossover operator to propagate descendants [9-11].

In the system made up of Web services, if key services suffer network attack or a service may be not available or some services need to be upgraded, a new service would be prepared to replace the old one. At this time, we should consider a range of issues. For example, whether new services can achieve all the functions of existing services for the system? whether new services would result in exclusions such as deadlock or other non-standard receiver when they interact with other services in the system? Therefore, services replacement should satisfy the following conditions [1-3]:

The functional properties of existing service are a subset of new service, which means that new service can provide the full functionality of existing service.

New services must maintain the balance of the existing system, which means the new service should ensure that the dependences between other services in the system and the old service remain unchanged.

Known from the requirements above, under condition of several available new services, the core issue of Web services replacing lies in "transparent replacement", which is to select a new service which has minimal impact on the existing dependencies of the services system to ensure that the vibration caused by replacing to the services system would be minimal.

2 Service Replacement Principles

2.1 Replacement Principles Based on QoS

If a service become invalid or its response time exceeds a certain limit, UDDI Registry will be retrieved to search for some candidate services with the expected functions according to two following principles.

1) Input/Output Matching: when a service is registered, its input/output set has been extracted by Provider Agent to from a service function matching table in which a registered item represents an abstract service and the same services are on behalf of services with same features and interface;

2) QoS Matching: the appropriate services can be selected by using some indicators such as response time, cost, availability, reliability and reputation, which means that all the cost indexes are smaller than the original service as well as all the efficiency indicators are bigger than the original service. In practice, we only need to find the service whose QoS indicators values $F(S_i)$ is bigger than the original service.

The non-functional properties of services are divided into two types - effectiveness and cost, so the QoS indicators values $F(S_i)$ of service can be figured out by using the equation 1.

$$F(S_i) = \sum_{j=1}^{n} q_j \times \omega_j \tag{1}$$

Where q_j is the value of each QoS index and ω_j is the weight of each QoS index.

In the process of service replacing, the service with biggest $F(S_i)$ in the candidate services set should be elected to replace the original service.

2.2 Reuse Principles

If the replacement of a candidate service is not better than the original one, the alternate implementation program of the composite service will be retrieved to update the QoS indexes of the used services and the superior degree of QoS indicators values will be by the equation 2 in order to select the service with the biggest $f(X_i)$ to replace the original service.

$$f(X_i) = \sum_{j=1}^{n} \frac{\left(q_j^n - q_j^0\right)}{q_j^0} \times \omega_j \tag{2}$$

Where q_j^0 and q_j^n are the index values of QoS attributes of original service and replacement service.

If $f(X_i) \geq 0$, the new combinations implementation program is relatively superior to the original proposal, and the new service can be used to replace the original service; else we should do a re-choose.

In summary, the design of service replacement algorithm must take into account the following rules:

1) After the existing service is replaced, the service composition model must still be well-formed;

2) The services which are represented by adjacent nodes in the graph are associated with a higher semantics degree, which means they can be replaced as a whole with a larger probability, so the service replacement algorithm should generate a connected sub-graph that contains the target service.
3) The number of all the connected sub-graphs is exponential and service combinations on behalf of most connected sub-graphs do not make sense in the semantic, so the target service should be generated based on the composite structure in order to improve the effectiveness of algorithm;
4) More services need to be replaced, the higher the switching costs, so the service replacement algorithm should gradually bottom-up expand the number of replacement services from the target service.

3 Replace Degree of Services

Web services system consists of a combination of several services, in which individual services are involved in a variety of dependencies. Hence before services replacing, we should do systematic analysis and classification to the dependencies of services to be replaced in order to provide data support of further calculations on the replace degree.

3.1 Dependency

In the service replacement process, the web services which need to be replaced may locate in a number of places in the services system, even indeed affect the whole system sometime. Because there is a wide variety of dependencies between various services in Web services system generally and these dependencies could transfer changes, the location of the extent of Web services is rather difficult. As we known, the first step of services replacing is to sort out dependencies between services in the system.

Dependency is a specific semantic connection existing between model elements, in which one element would be changed with the change of elements that depend on to ensure the integrity and consistency of the system. According to the definition, services $WSSource$ depend on the model elements $WST\arg et$ by dependency type $DType$, recorded $Dep = (WSSource, WST\arg et, DType)$, where dependency type $DType$ contains seven types such as interface dependence, data dependence, control dependence, timing dependence, state dependence, causal dependence, I/O dependence [8].

3.2 Replace Degree

If the new service could remain the compatibility relations of the old service with other services unchanged, we say that it can replace the old service. It can be seen that the dependence of the old service is inversely proportional to its replace degree, which means the higher the dependence degree of replacement services in the service system, the lower the replace degree.

Definition 1. Web services could be expressed as a six-tuple $WS = (\sum_{in}, \sum_{out}, ST, S_0, S_T, \Delta)$, where $\sum_{in}$ is the input set, $\sum_{out}$ is the output set, ST is the state set, S_0 is the initial state, S_T is the final state and Δ is the state transition function $\Delta : ST \times (\sum_{in} \cup \sum_{out}) \to ST$.

Definition 2. Replace degree indicates the depth and breadth of influence of dependencies between replace service and other services on replacement process and its results, which is represented as $\mathrm{Re}\,Deg = (WSSource, WST\arg et, RDepth, RScope)$ in this paper where $RDepth$ is the depth of replacement influence as well as $RScope$ is the breadth of replacement influence.

The formula of Replace degree is as follows:

$$\mathrm{Re}\,Deg = \frac{1}{\sum_{i=1}^{n} Dep_i} \tag{3}$$

Where n is the number of services in the actual impact set.

Seen from the formal expression of dependence, dependence and influence spread is opposite, shown as the figure 1.

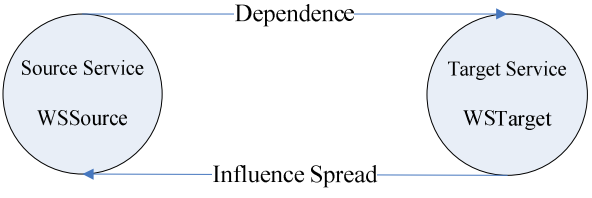


Fig. 1. Dependence and Influence Spread

Known form the figure 1, the direction of dependence Dep is from source service $WSSource$ to the target service $WST\arg et$, which means $WSSource$ depends on $WST\arg et$; the direction of influence spread is from the target service $WST\arg et$ to source service $WSSource$, which means $WSSource$ should be changed when $WST\arg et$ is changed to guarantee the semantically correct and structural integrity of the entire service system.

Therefore, when we analyze influence spread and dependencies of service replacing, we use complex network structure based on the directed graph.

4 Replacement Algorithm Based on Complex Networks

In this paper, the topology of services system is abstracted into a graph consisted by point set and edge set. That is to say, web services in services system are seen as points in a graph as well as their dependencies are seen as edges in a graph, shown as the figure 2. By this way, the graph made up of these points and edges could reflect the map from services system to complex network. At this time, we get our Service Replacement Algorithm using Complex Network (SRA-CN).

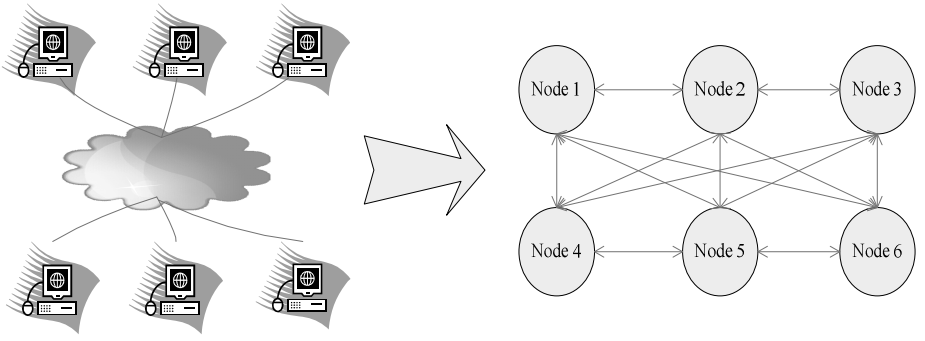


Fig. 2. Map from Services System to Complex Network

In the directed network in the figure 2, node degrees are classified into in-degree and out-degree [4-6]:

In-degree is the number of edges from other points to the point. The larger in-degree of a point, the stronger its dependence.

Out-degree is the number of edges from the point to other points. The larger out-degree of a point, more dependent the point to other points.

Therefore, we use in-degree and out-degree of a point in complex network to express the strength of dependencies between services in services system, and give its formula as follow:

$$Dep_i = Deg_{in} + Deg_{out} \tag{4}$$

Where i is the point for the old service to be replaced, Deg_{in} is the in-degree of point i and Deg_{out} is the out-degree of point i.

Based on the description above, we get the specific process of SRA-CN as follows.

Algorithm: Service Replacement Algorithm using Complex Network (SRA-CN)
Input: new services set WS={WSre, WSre, WSre,, …WSn} for service to be replaced WSre, dependence degree Dep_i, replace degree $\mathrm{Re}\,Deg$, transmission probability $P_i(j)$
Output: combination service WSc
Description:

1. if $\mathrm{Re}\,Deg \geq 0.6$
// which means the service to be replaced is a key point in the complex network.
2. print {"The service cannot be replaced!"}
3. else While $WS \neq \phi$ do
4. for each $WS_i \in WS$
5. compute Dep_i using (2)
6. $WS_N = Max(Dep_i)$
// The service whose dependence degree with related services of the service to be replaced is largest would be selected to be the new service.
7. end if
8. return WSc // put out the combination service

Based on the SRA-CN, we get Web service composition algorithm based on SRA-CN further as follows:

```
Algorithm: Service Composition Algorithm based on SRA-CN
Input: combination service WSC1 (X) with abnormal performance
Output: replaced combination service WSC1 (X)
Description:
    1. if  a service become abnormal
     // which means there is a service needed to be replaced.
    2. Sj=SRA-CN•S•
                 f(Xi) ≥ 0
    3.   if
     // which means the new service is superior to the original service.
    4.    else print{"A suitable candidate is not found!"}
    5. end if
    6.  return WSC2(X)
```

5 Performance Analysis

In order to evaluate the performance of SRA-CN, we designed comparative experiments with the algorithm in [7].

Experimental parameters are 300 web services which belonged to 50 classes and randomly distributed in 300 nodes in the network. We executed services combination 100 times, and set randomly 10 failure occurring points in the process of services replacing with transmission probability 0.6.

The performances of an algorithm mainly contain efficiency (that is the response time of algorithm to service replacing) and replacement time (that is the time of algorithm to complete service replacement), and the results are shown in the figure 3 and 4.

The Figure 3 compared the replacing response time (ms) with different number of services combination.

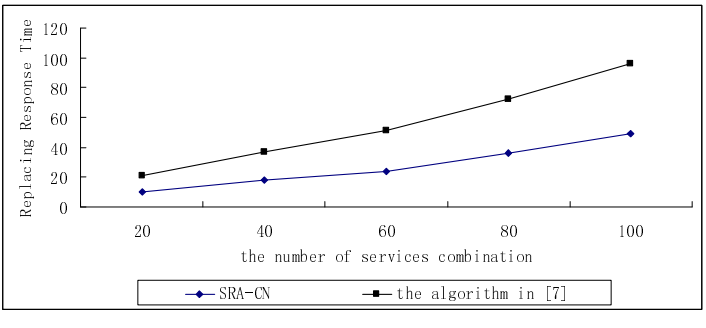


Fig. 3. Comparative Analysis of Replacing Response Time

The Figure 4 compared the replacing completion time (ms) with different failure occurring points.

It can be seen that the running time of SRA-CN was longer when running services were less and the running time of SRA-CN was shorter than the algorithm in [7] when running services were more. That is to say, SRA-CN has a better efficiency and availability.

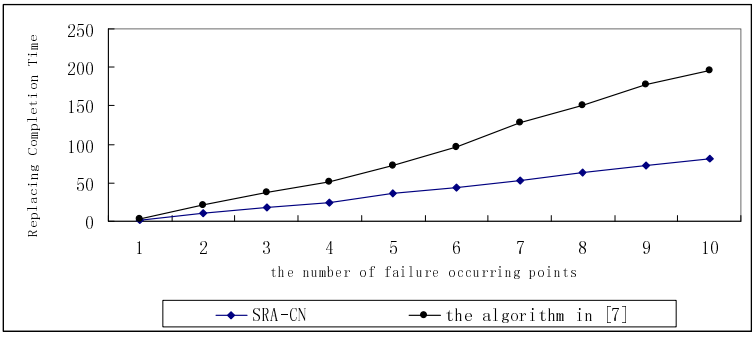


Fig. 4. Comparative Analysis of Replacing Completion Time

6 Service Replacement Implementation Architecture Based on SRA-CN

We use SRA-CN in the actual Web service replacement by using the replacement implementation architecture shown in the Figure 5 in which WS-BPEL engine and SOAP engine are expanded to reduce the coupling between modules and improve the scalability and flexibility of the engines.

Known from the Figure 5, the replacement implementation architecture based on SRA-CN is made up of three modules.

1) Service Replacement Module (SRM): SRM is used to maintain a collection of candidate services that provide the same or similar functions. In the SRM, the assessed value of candidate services will be got to select the best service to meet customer needs, and all relevant information of services will be recorded in the replacement process;
2) Service Evaluation Module (SEM): SEM provides external interfaces to receive various users needs of services, returns evaluation results of candidate services, and updates a variety of related data used by SRA-CN;
3) Monitoring Components (MC): MC firstly gets the information of invoking service endpoint such as address and name, secondly start the timer for timing the binding, calling & replacing of web service, and finally stores related information of services from replacement.

Under this architecture, the new service should do local screening and global screening before replacement.

The first step is to do a local screening to the replacement service by using actual historical record of the web service:

1) to collect the historical data of the original service, including response time, reliability & cost;
2) to calculate the dependencies and replace degrees of the candidate services;

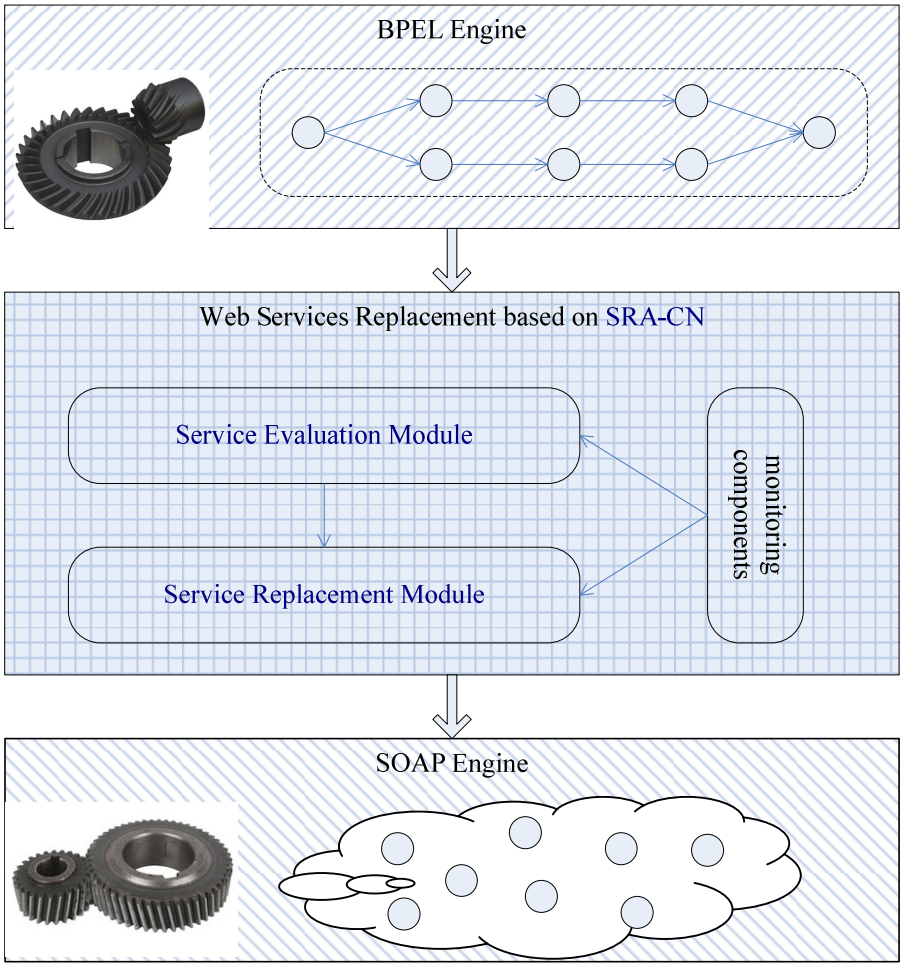


Fig. 5. Web Services Replacement Architecture Based on SRA-CN

3) According to service replacement principles and practical experience, candidate services whose replace degrees has been calculated will be weighted with a relaxation factor to avoid exclude some effective alternative services.

By following three steps above, the candidate services can be filtered locally. As for the candidate services which meet local constraints, we need a second round of global screening for them:

1) The candidate services which pass the local screening will be replaced into the combination services, and the dependencies of each service in the combination services will be recalculated;

2) The dependencies of combination services will be compared to makes sure that all of them meet the global constraints, and non-compliant alternative service should be excluded.

After the screening of candidate services, we can speak them into the service evaluation module and services replacement module to finish the service replacing under the guidance of monitor component.

7 Conclusion

Because the features of services combination such as correlation, select the best and evolving fit self-organizing & self-evolution of complex network, we propose a services replacement algorithm using complex network SRA-CN. In the SRA-CN, complex network is used to describe & calculate dependencies between web services and two calculation factors, replace degree and transmission probability, are designed to Measure the impact of dependence on service replacing. At the end, we put our algorithm and another similar algorithm in [7] together to verify the performance of SRA-CN by two comparative experiments.

Based on the work we have done, our next research focus is to design dynamic service replacement algorithm using complex network. We plan to add time factor into SRA-CN to reflect the actual process of services replacement more comprehensive, and hope to do more in-depth study in the field of adaptive Web services combination.

References

1. Zhang, Y., Fang, B., Xu, C.: Preference Ontology-Oriented Metric Model For Trustworthy Web Services. International Journal of Intelligent Systems 26(2), 158–168 (2011)
2. Zhang, Y., Fang, B., Xu, C.: Trustworthiness_Metrics Model For Internetware. In: International Symposium on Education and Computer Science, pp. 326–329 (2009)
3. Zhang, L., Zhang, J., Hong, C.: Services Computing. Tsinghua University Press, Beijing (2007)
4. Cai, W., Zhao, H., Zhang, H.H.: Static Structural Complexity Metrics For Large-scale Software. Special Issue on Software Engineering and Complex Networks of Dynamics of Continuous, Discrete and Impulsive Systems Series B 14(S6), 12–17 (2007)
5. Zheng, X.L., Daniel, Z., Li, H.Q.: Analyzing Open-source Software Systems As Complex Networks. Physica A: Statistical Mechanics and Its Applications 387(24), 6190–6200 (2008)
6. Jenkins, S., Kirk, S.R.: Software Architecture Graphs As Complex Networks: A Novel Partitioning Scheme to Measure Stability and Evolution. Information Sciences 177(12), 2587–2601 (2007)
7. Zaremba, M., Migdal, J., Hauswirth, M.: Discovery of Optimized Web Service Configurations Using A Hybrid Semantic and Statistical Approach. In: Proceedings of the 7th IEEE International Conference on Web Services, Los Angeles, USA, pp. 149–156 (2009)

8. Ma, L., Wang, H., Lu, Y.: The Design of Dependency Relationships Matrix To Improve The Testability of Component-based Software Quality Software. In: 6th International Conference on Volume (QSIC 2006), pp. 93–98 (2006)
9. Chang, M., Wu, S., Heh, J.-S.: Making the Real World as a Game World to Learners by Applying Game-Based Learning Scenes into Ubiquitous Learning Environment. In: Pan, Z., Cheok, D.A.D., Müller, W., El Rhalibi, A. (eds.) Transactions on Edutainment I. LNCS, vol. 5080, pp. 261–277. Springer, Heidelberg (2008)
10. Sauvé, L.: Design Tools for Online Educational Games: Concept and Application. In: Pan, Z., Cheok, A.D., Müller, W., Rhalibi, A.E. (eds.) Transactions on Edutainment II. LNCS, vol. 5660, pp. 187–202. Springer, Heidelberg (2009)
11. Ahn, S., Shi, C.-K.: Exploring Movie Recommendation System Using Cultural Metadata. In: Pan, Z., Cheok, A.D., Müller, W., Rhalibi, A.E. (eds.) Transactions on Edutainment II. LNCS, vol. 5660, pp. 119–134. Springer, Heidelberg (2009)
12. Lazovik, A., Aiello, M., Papazoglou, M.: Planning and monitoring the execution of Web service requests. Journal on Digital Libraries, 235–246 (2006)
13. Erdos, P., Renyi, A.: On the Evolution of Random Graphs. Publ. Math. Inst. Hung. Acad. Sci. (S1103-467x), 17–60 (1960)
14. Watts, D., Strogatz, S.: Collective Dynamics of "Small World" Networks. Nature (S0028-0836), 440–442 (1998)
15. Newman, M.J., Watts, D.: Renormalization Group Analysis of the Small-World Network Model. Phys. Lett. A (S0375-960), 41–346 (1999)
16. Barabási, A., Albert, R.: Emergence of Scaling in Random Networks. Science (S0036-8075), 509–512 (1999)
17. Albert, R., Jeong, H., Barabási, A.: Diameter of the World Wide Web. Nature (S0028-0836), 130–131 (1999)

A Virtual Training System Using a Force Feedback Haptic Device for Oral Implantology

Xiaojun Chen[1], Yanping Lin[1], Chengtao Wang[1],
Guofang Shen[2], and Xudong Wang[2]

[1] Institute of Biomedical Manufacturing and Life Quality Engineering,
State Key Lab of Mechanical System and Vibration, School of Mechanical Engineering,
Shanghai Jiao Tong University, Shanghai, China
[2] Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine,
Shanghai, China
xiaojunchen@163.com

Abstract. As an integral part of dental practice, oral implant therapy has been accepted worldwide and become increasing important over the past few years. However, the placement of implants is not without risk due to anatomically complex operation sites in the cranio-maxillofacial region. Currently, there is a trend towards computer-aided implant surgery. In this study, a comprehensive preoperative planning and virtual training system is developed on the basis of the force feedback haptic device (Omega.6), immersive workbench (Display 300), and the software toolkit of CHAI3D. With the use of this system, the resulting data of the preoperative planning can be transferred, and surgical simulation of the plan can be vividly realized. In this way, the surgeon can grasp the feeling of osteotomy procedure, gain experience and therefore improve his skills during the actual dental implant surgery. This pilot study proves helpful for the inexperienced surgeons; however, more clinical cases will be conducted to demonstrate its feasibility and reliability.

Keywords: surgical planning, virtual training, force feedback haptic device, dental implant surgery.

1 Introduction

During the past decades, as a welcome alternative to ill-fitting dentures or bridgework, dental implants are frequently the best treatment option for replacing missing teeth. There are many advantages of dental implants over fixed bridges. They can be fixed with no change to the overall structure of the face or jawbone, as is often the case when a tooth is removed. Therefore, dental implants are now becoming a more popular treatment to replace missing teeth, as they provide a longer-term solution, slow down bone loss and preserve nearby healthy tooth tissue. However, dental implant surgery is not without risk. Prior to commencement of surgery, careful and detailed planning is required to identify vital structures such as the inferior alveolar nerve or the sinus, as well as the shape and dimensions of the bone to

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 232–240, 2012.
© Springer-Verlag Berlin Heidelberg 2012

properly orient the implants for the most predictable outcome [1, 2]. Although several CAD/CAM systems for preoperative planning and the fabrication of surgical guides have been developed commercially available, such as SimPlant(Materialise, Leuven, Belgium), NobelGuide (Nobel Biocare, Yorba Linda, CA), ImplantMaster(I-Dent Imaging, Ft. Lauderdale, FL), etc.[3,4], they do not help inexperienced surgeons to grasp the feel of osteotomy procedure during an actual operation.

With the development of computer science, there is a trend towards of simulating surgery procedure based on Virtual Reality Technique, which helps surgeons improve surgery plans and practice surgery process on 3D patient specific models [5]. The simulator surgery results can be evaluated before the surgery is carried out on real patient. So, when the real surgery takes place, the surgeon is already familiar with all the specific operations that are to be employed, therefore, his surgical skills can be greatly improved and higher predictability of the treatment outcome can be realized. Currently, virtual surgical training technique focuses on laparoscopic surgery, and several surgical simulators are commercially available in this field, such as SIMENDO Laparoscopy (SIMENDO, Netherlands), medical (Immersion inc , USA), LapSim (Surgical Science Sweden AB, Göteborg, Sweden), LAP Mentor (Simbionix Ltd., Lod, Israel), Mentice MIST (Mentice AB, Göteborg, Sweden),etc[5]. The surgical simulator technology has already been tested and proved by many research groups and clinical physicians in numerous applications[6,7]. However, its applications in dental implant surgery are rarely reported.

This study presents an integrated surgical planning and virtual training system that encompasses the 3D medical modeling, preoperative surgical planning, and virtual training. With the use of this system, the surgeon can identify vital anatomical structures, grasp the feeling of osteotomy procedure, gain experience and therefore improve the surgical skills during dental implant surgery.

2 Material and Methods

2.1 The Framework and Development of the Preoperative Planning Software

The framework of the system is described as follows: A modular software named CAPPOIS (Computer Assisted Preoperative Planning for Oral Implant Surgery) was developed for preoperative planning, including medical image importing, image segmentation and 3D-Reconstruction, 2D/3D geometrical measurements, optimization design of the position and orientation of oral implants, bone density analysis, bone volume measurement for maxillary defects, simulation of tumor resection and free bone graft, etc. Then, the resulting data of the preoperative planning can be transferred to a self-developed virtual training system using a force feedback haptic device. Based on this training system, the surgeon can be familiar with the preoperative plan and experience the specific operations, especially the osteotomy procedure, before the surgery.

Various algorithms in the field of computer graphics and medical image processing are involved in CAPPOIS, including DICOM file parsing, image segmentation and

3D-visualization [8], surface decimation [9], cutting, spatial search and 3D distance computing [10], volume measurement, spline curve generation, multi(curved)-planar reconstruction, registration, etc. For each of these algorithms, we have developed a set of dynamic link libraries (DLL) using Microsoft Visual C++, as well as the Visualization Toolkit (VTK, an open source, freely available software system for 3D computer graphics, image processing, and visualization etc., http://www.vtk.org/) and Insight Toolkit (ITK, an open-source software toolkit for performing registration and segmentation, http://www.itk.org/) via object oriented programming methodology. This basis can be extended by virtually any new approach or algorithm, which then becomes seamlessly integrated into the method set of the software framework. The aim is to provide well-defined levels of abstraction (the hiding of implementation details) from the individual components, so that new technology can be incorporated into the system without a complete software rewrite. As for graphical user interfaces (GUI), we choose QT (a free, open source, and cross-platform application development framework widely used for the development of GUI programs, http://qt.nokia.com/products).

Since these well-known open-source toolkits such as VTK, ITK, and QT are involved, the functions of CAPPOIS can be easily extended with the support of the new emerging, leading-edge algorithms and technologies in the field of the medical image computing, such as GPU-accelerated volume rendering, automatic functionality segmentation using the level set method, efficient non-rigid registration for multi-modality imaging, etc. Using CAPPOIS, prosthodontists can select a certain type of virtual implants and place them into the ideal areas according to the information concerning the relevant anatomical structures. The virtual implants will be rendered on all of the 2D/3D views; therefore, the type, number, size, position, and orientation of the implants can be determined explicitly by iterative optimization taking into account prosthetic requirements and available local bone. The examples of CAPPOIS for preoperative planning of the implant placement in the mandible and maxilla are respectively shown in Fig.1. and Fig.2.

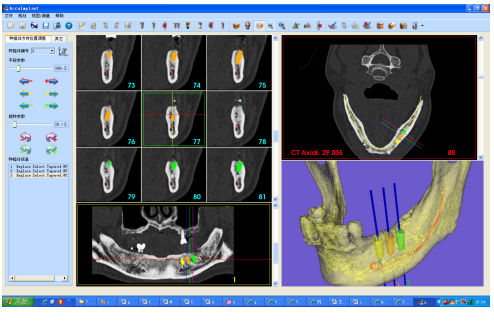


Fig. 1. An example of preoperative planning for the implant placement in the mandible with the use of CAPPOIS

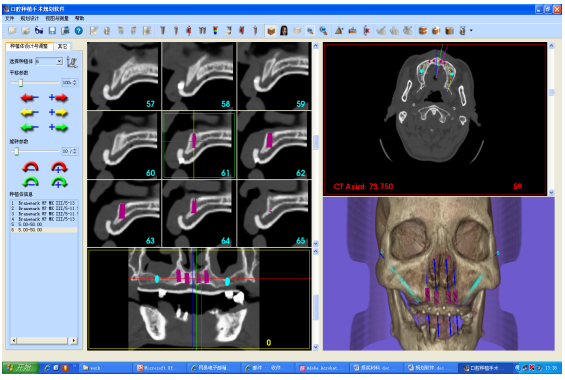


Fig. 2. An example of preoperative planning for the implant placement in the maxilla with the use of CAPPOIS

2.2 The Development and Application of the Virtual Surgical Training System

The virtual surgical training system is based on the force feedback haptic device (Omega.6, Force Dimension, Switzerland) and immersive workbench shown in Fig. 6. The specifications of the haptic device are as follows: workspace: ⌀160 x L 110 mm; maximum force to be sensed: 12.0N; resolution: linear, < 0.01 mm ; interface: USB 2.0 [11]. As for the immersive workbench, the SenseGraphics Display Series 300 is used, since it is compatible with a range of commercial available haptic hardware, such as SensAble Omni, SensAble dual Premium, ForceDimension Omega, etc. It is featured with a "Ikea-knock down solution" frame with silver-coated mirror, so that true 3D stereo performance in high resolution can be achieved with shutter glasses and a high performance 120 Hz LCD-monitor[12]. As for the development of the software, we chose the toolkit of CHAI3D [13], which combines an extensive number of force rendering algorithms, including the finger-proxy model, potential fields and implicit based models which allow programmers to easily develop sophisticated simulations with integrated force-feedback capabilities. Since the CHAI3D can offer all the basic functions ranging from position reading to force programming in Cartesian space, it can easily be extended to incorporate more advanced specific features, such as the surface deformation of the 3D tissue model, the simulation of the force feedback during virtual drilling of the bone, etc. On the other hand, modeling of surgical instruments can be easily realized via 3D feature-based solid modeling software (such as Pro/E, UG, etc.) according to the geometric information of the instrument.

After the modeling process, collision detection can be achieved through bounding-box based algorithms. There are various kinds of bounding boxes developed for many applications in computer graphics and visualization[24], such as OBB (oriented bounding box)[19-21], k-DOPs (Discrete Orientation Polytope)[22], AABBs (Axis-aligned bounding boxes)[23], etc. Compared with other kinds of bounding boxes, although AABBs may be particularly poor approximations of the set that they bound, leaving large “empty corners”, they are simple to compute and they allow for very efficient overlap queries[22]. Since real-time collision detection is of critical

importance for surgery simulation, e.g., haptic force-feedback can require on the order of 1,000 intersection queries per second[20], AABBs are adopted in this study. AABBs for the cranio-maxillofacial model and surgical instrument models are constructed top-down, by recursive subdivision. At each recursion step, the smallest AABB of the set of primitives is computed, and the set is split by ordering the primitives with respect to a well-chosen partitioning plane. This process continues until each subset contains one element. Meanwhile, the aligned bounding boxes of the surgical instrument model will be realigned in real time as they become unaligned due to unavoidable rotation during the motion. In this way, the efficient collision detection between the cranio-maxillofacial model and surgical instrument models can be realized.

So, after the surgeon imports the resulting data of the preoperative planning based on CAPPOIS, he can manipulate the handle of the force feedback haptic device to simulate the movement of the virtual dental handpiece. When the virtual handpiece model reaches the surface of jawbone, the surgeon can sense a feeling of object collision (Shown in Fig.4). When he continues to drill the bone, the amount of the reactive force varies with the anatomical structure which the hand piece reaches (For example, when the tip of the handpiece reaches the cortical bones, the feedback force is large, while it enters cancellous bones, the force is relatively far more smaller). Meanwhile, during the osteotomy procedure, the virtual cavities in the jawbone are formed in real time. With the use of the immersive workbench, the image can be viewed through 3-D glasses which give a realistic sense of depth (Shown in Fig.5). In this way, the trainees can begin to develop skills at an earlier stage which is likely to be of fundamental importance in the future.

Currently, we are colleting the data concerning the actual interactive force for the activities such as cutting, drilling, piercing, etc., during the oral and maxillofacial surgery. After that, we will investigate the related factors (such as the bone density, the speed of the surgical instrument, the cutting depth, etc.), which affect the cutting force during the surgery, aiming at building a regression equation to predict the cutting force according to the statistical analysis. If it is obtained, it will be applied to the current virtual training system, so that more accurate feedback force can be achieved.

Fig. 3. The virtual reality force feedback haptic device (Omega.6, Force Dimension) and immersive workbench(Display Series 300, SenseGraphics)

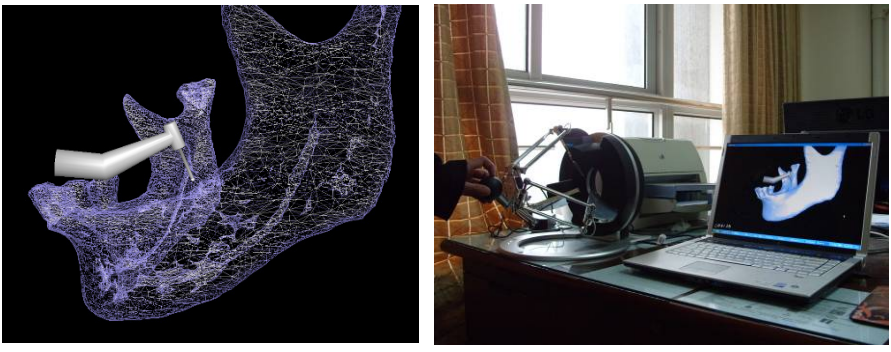


Fig. 4. The experiment of the virtual surgical training system with the force feedback haptic device

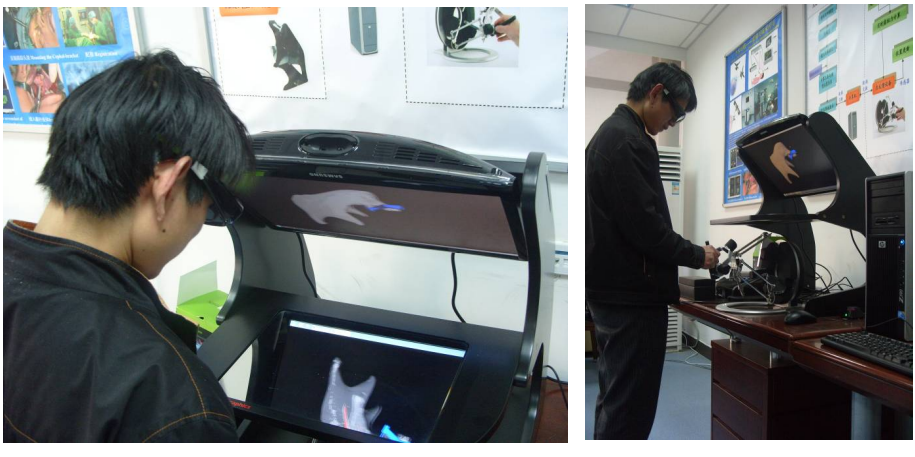


Fig. 5. The experiment of the virtual surgical training system with the force feedback haptic device and immersive workbench

3 Discussion and Conclusion

While the advent of medical imaging systems has added "X-ray" vision to the surgeon's armament, allowing him to visualize tissue color, perfusion, and texture, it has largely deprived him of the sense of feel that open surgery provides. The surgeon's sense of touch – his haptic sense – enables him to identify otherwise obscure tissue boundaries, to differentiate normal and abnormal tissue, and to determine the appropriate amount of force to use in cutting, clamping, and suturing[6,14]. Therefore, in terms of improving the surgical skills, the virtual surgical training system, which allows the surgeons to feel the texture of organs and even manipulate them via surgery simulation, are very helpful, especially for the

inexperienced surgeons. Currently, virtual surgical training systems are widely used in laparoscopic surgery[7].

However, there are only a few reports about surgery simulation for oral implantology. For example, using the Microsoft® Sidewinder™ force feedback joystick, Neumann, P. ,et al., developed virtual tools available for the surgeon to carry out a three~dimensional (3D) cephalometrical analysis and interactively define bone segments from skull and jaw bones[15]. Xia, J., et al, present a PC-based virtual reality surgical simulation for orthognathic surgery (VRSSOS), for surgeons to perform virtual orthognathic surgical planning in three dimensions[16]. Schendel, S., et al, developed a computer-based surgical simulation system, which allows the user to interact with a virtual patient to perform the traditional steps of cleft-lip repair[17]. Kusumoto, N., et al, developed a novel support system for oral implant surgery, which involves manipulating a 3D CT image of a jawbone with a virtual reality force feedback haptic device. Through this virtual system, the haptic experience of bone drilling with vibration and the sound of the contra-angle handpiece could be realized[18].

Compared with those above-mentioned studies, the virtual training system for oral implantology developed in this study are more comprehensive with more accurate force feedback haptic device(Omega.6) and immersive workbench(Display 300). A modular software named CAPPOIS (Computer Assisted Preoperative Planning for Oral Implant Surgery) was developed for preoperative planning. With the introduction of DLL (Dynamic Link Libraries) and some well-known free, open source software libraries such as VTK, ITK, and QT, a plug-in evolutive software architecture was established, allowing for expandability, accessibility, and maintainability in our system. Then, the following virtual surgical training system is developed based on the force feedback haptic device, i.e., Omega.6, and the software toolkit of CHAI3D. With the use of this system, the resulting data of the preoperative planning using CAPPOIS can be imported, and the surgeon can "perform surgery" upon the 3D virtual jawbones by manipulating the virtual handpiece, while the Omega.6 simultaneously provides realistic force-feedback and collision detection during the osteotomy procedure. With the use of the immersive workbench (Display 300), more realistic and 3D stereo sense during the whole procedure can be felt by the surgeon when being surrounded in an engrossing total virtual environment.

Nevertheless, this pilot study proves that this system is helpful for the inexperienced surgeons to grasp the feel of osteotomy procedure during dental implant surgery. However, more clinical cases will be conducted to demonstrate its feasibility and reliability, and experiments will be performed, aiming at providing more accurate feedback force.

Acknowledgments. This project was supported by Natural Science Foundation of China (Grant No.:51005156, 81171429), the National High Technology Research and Development Program (Grant No.: SQ2009AA04ZX1485930). The authors would also like to thank Ms. Catherine Liu for reviewing the manuscripts and offering constructive criticisms, which were incorporated in the final form.

References

1. Askary, A.S., Meffert, R.M., Griffin, T.: Why do dental implants fail? Part I. Implant Dentistry 8, 173–185 (1999)
2. Askary, A.S., Meffert, R.M., Griffin, T.: Why do dental implants fail? Part II. Implant Dentistry 8, 265–277 (1999)
3. Lal, K., White, G.S., Morea, D.N., Wright, R.F.: Use of stereolithographic templates for surgical and prosthodontic implant planning and placement. Part I. The Concept. Journal of Prosthodontics 15, 51–58 (2006)
4. Lal, K., White, G.S., Morea, D.N., Wright, R.F.: Use of stereolithographic templates for surgical and prosthodontic implant planning and placement. Part II. A Clinical Report. Journal of Prosthodontics 15, 117–122 (2006)
5. http://en.wikipedia.org/wiki/Virtual_surgery
6. Seymour, N.E., Gallagher, A.G., Roman, S.A., O'Brien, M.K., Bansal, V.K., Andersen, D.K., Satava, R.M.: Virtual reality training improves operating room performance - Results of a randomized, double-blinded study. Annals of Surgery 236, 458–464 (2002)
7. Grantcharov, T.P., Kristiansen, V.B., Bendix, J., Bardram, L., Rosenberg, J., Funch-Jensen, P.: Randomized clinical trial of virtual reality simulation for laparoscopic skills training. British Journal of Surgery 91, 146–150 (2004)
8. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3D surface construction algorithm. ACM Computer Graphics 21, 38–44 (1987)
9. Schroeder, W.J., Zarge, J.A., Lorensen, W.E.: Decimation of triangle meshes. Proceedings ACM Siggraph 26, 65–70 (1992)
10. Schroeder, W., Martin, K., Lorensen, W.: The Visualization Toolkit: An Object-Oriented Approach To 3D Graphics. Prentice Hall (1997)
11. http://www.forcedimension.com/omega3-specifications
12. http://www.sensegraphics.com/index.php?page=shop.product_details&category_id=7&flypage=shop.flypage_sensegraphics&product_id=45&option=com_virtuemart&Itemid=83
13. http://www.chai3d.org/
14. http://www.mediligence.com/rpt/rpt-s165.htm
15. Neumann, P., Siebert, D., Schulz, A., Faulkner, G., Krauss, M., Tolxdorff, T.: Using virtual reality techniques in maxillofacial surgery planning. Virtual Reality 4, 213–222 (1999)
16. Xia, J., Samman, N., Chua, C.K., Yeung, R.W.K., Wang, D., Shen, S.G., Ip, H.H.-S., Tideman, H.: PC-based Virtual Reality Surgical Simulation for Orthognathic Surgery. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) MICCAI 2000. LNCS, vol. 1935, pp. 1019–1028. Springer, Heidelberg (2000)
17. Schendel, S., Montgomery, K., Sorokin, A., Lionettia, G.: Surgical simulator for planning and performing repair of cleft lips. Journal of Cranio-Maxillofacial Surgery 33, 223–228 (2005)
18. Kusumoto, N., Sohmura, T., Yamada, S., Wakabayashi, K., Nakamura, T., Yatani, H.: Application of virtual reality force feedback haptic device for oral implant surgery. Clinical Oral Implants Research 17, 708–713 (2006)
19. Ding, S., Mannan, M.A., Poo, A.N.: Oriented bounding box and octree based global interference detection in 5-axis machining of free-form surfaces. Computer-Aided Design 36, 1281–1294 (2004)

20. Gregory, A., Lin, M.C., Gottschalk, S., Taylor, R.: Fast and accurate collision detection for haptic interaction using a three degree-of-freedom force-feedback device. Computational Geometry-Theory and Applications 15, 69–89 (2000)
21. Chang, C.T., Gorissen, B., Melchior, S.: Fast Oriented Bounding Box Optimization on the Rotation Group SO(3, R). ACM Transactions on Graphics 30(5), Article 122 (2011)
22. Klosowski, J., Held, M., Mitchell, J., Sowizral, H., Zikan, K.: Efficient Collision Detection Using Bounding Volume Hierarchies of k-DOPs. IEEE Transactions on Visualization and Computer Graphics 4, 21–36 (1998)
23. Zhang, X., Kim, Y.: Interactive collision detection for deformable models using streaming AABBs. IEEE Transactions on Visualization and Computer Graphics 13, 318–329 (2007)
24. Agarwal, P.K., De Berg, M., Gudmundsson, J., Hammar, M., Haverkort, H.J.: Box-trees and R-trees with near-optimal query time. Discrete & Computational Geometry 28, 291–312 (2002)

Modeling and Analysis of the Business Process of the Supply Chain of Virtual Enterprises

Min Lu and Haibo Zhao*

Lishui University, Lishui, 323000, Zhejiang Province, China
zjlszjz@163.com, zhbbu@sina.com

Abstract. In this paper, UML and object-oriented Petri net are combined to model the distributed business process of the supply chain of virtual enterprises. First the virtual enterprise global business process model is established by using UML sequence diagram. Then the previously established model is converted to Petri net model. Finally emulation tool Hpsim of Petri net is applied to carry out emulation operation to analyze the reasonability of the Petri net model.

Keywords: supply chain of virtual enterprises, business process, Petri net, emulation tool.

1 Introduction

A virtual enterprise (VE) is an interim union of enterprises composed of various member enterprises, with the purposes of promoting the sharing of techniques, core abilities and resources under current market opportunities and enhancing the cooperation between enterprises. Supply Chain Management (SCM) is implemented in a virtual enterprise with the hope of realizing the seamless convergence of the overall dynamic union of the virtual enterprise and speeding up enterprises' reacting to market demands, thus to seize opportunities and gain a larger share of the market. A virtual enterprise is composed of many enterprises of different nature and there are different suppliers, manufacturers, distributors and so on in the supply chain of a virtual enterprise, who play different roles in the supply chain. When designing their enterprise system, these enterprises in nodes of the supply chain only pay attention to the production process rather than to other factors that may affect the enterprise's competitiveness and the overall supply chain, which results in seam in the business process of inter-enterprise cooperation, affecting the running of the virtual enterprise, lowering the operating efficiency of the supply chain of the virtual enterprise, making it impossible to effectively complete market orders and even finally causing the running of the virtual enterprise a failure. Therefore, implementation of Supply Chain Management will inevitably involve the design of the business process of the supply chain.

The effective operation of the supply chain of a virtual enterprise requires the support of a highly efficient and concise business process. Modeling of the business

* Corresponding author.

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 241–249, 2012.
© Springer-Verlag Berlin Heidelberg 2012

process of the supply chain of a virtual enterprise is a kind of cross-organizational business process modeling. Models previously established are nearly all about the modeling of internal business process of enterprises. Some researches have set up cross-organizational business process models but the models are too big, which are difficult to analyze and may cause the system's state explosion. In addition, in the supply chain of a virtual enterprise, some member enterprises will try their best to maintain their own commercial secrets and not be willing to disclose detailed information about their internal sub-business processes, so it is quite impossible to get a detailed model of the distributed whole business process for a virtual enterprise. All those models previously established pay neither attention to the confidentiality of each enterprise nor to the business processes resulted from inter-enterprise cooperation, so the production, supply and sales links of enterprises have not formed a true "chain".

In order to establish a cross-enterprise-boundary business process model for virtual enterprises and solve the state explosion and confidentiality problems, this paper combines UML and object-oriented Petri net to establish a model of distributed business process for the supply chain of virtual enterprises and applies modular approach in the modeling process. The whole model is divided into whole business process (WBP) module and sub-business process (SBP) module, to solve such problems in previous models as large size and difficulties in analyzing.

2 Modeling of Whole Business Process of Virtual Enterprises (VE-WBP)

2.1 Interactive Model of VE-WBP Defined by UML Sequence Diagram

UML sequence diagram can describe the message exchange between objects, so it is applicable to describing the interactions among member enterprises in the distributed business process of a virtual enterprise. Figure 1 shows the application of the UML sequence diagram in describing the interaction scene in the whole business process of a virtual enterprise. This paper will start from this simplified interaction scene of virtual enterprises and put forward methods for modeling and analyzing the distributed business process of virtual enterprises.

The interactive scene of virtual enterprises in Figure 1 involves four objects: customer and three member enterprises of a virtual enterprise, VELeader, VEPartner1, and VEPartner2. The interactive process among the four objects are as follows: first, a customer makes a product order to the core enterprise VELeader (asynchronous message m1); second, having received the product order, VELeader gives part orders to both member enterprise VEPartner1 and member enterprise VEPartner2 (asynchronous messages m2 and m3); having received the orders, VEPartner1 and VEPartner2 should conduct synchronization of specific activities at certain times (synchronous message m4); then VEPartner1 and VEPartner2 deliver finished parts (asynchronous messages m5 and m6); finally, VELeader delivers the final products to the Customer (asynchronous message m7).

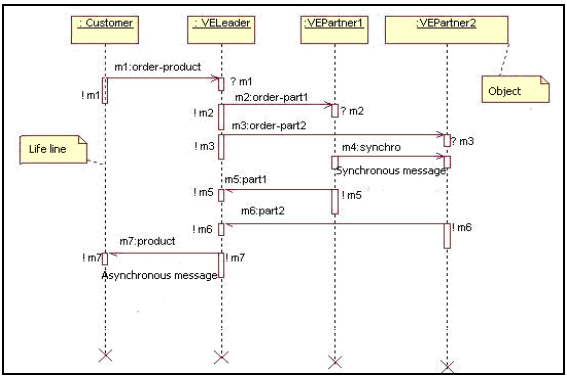


Fig. 1. The interactive model of UML sequence diagram

From above example, it can be seen that UML sequence diagram can easily show the interactions among various member enterprises and thus the interactive model of distributed VE-WBP (Virtual Enterprise-Whole Business Process) is established. The interactive model of VE-WBP, described by the UML sequence diagram, only contains the interface activities in which various member enterprises are interested but not the sub-business process actually executed by member enterprises. It also does not describe the actual way of executing the activities.

2.2 Transition from Model of UML Sequence Diagram to Model of Petri Net

In order to solve the problem that it is difficult to carry out analysis and verification by using the UML-based model of business process model of virtual enterprises, the model described by UML sequence diagram should be transited into a model described by Petri net. The UML sequence diagram model can be converted into the Petri net model through the following mapping Table 1.

Table 1. Mapping table

UML sequence diagram	Petri
Asynchronous message events	Asynchronous message transition
Asynchronous message	Asynchronous message place
Synchronous message events	Synchronous message transition
Synchronous message	No place (may not be represented by place)
Life line	Connection of transition and place

Through transition from UML sequence diagram to Petri net, the interactive model of Petri net of virtual enterprises, shown in Figure 2, is obtained. The interface messages about the business process subnets of member enterprises are also contained in this model.

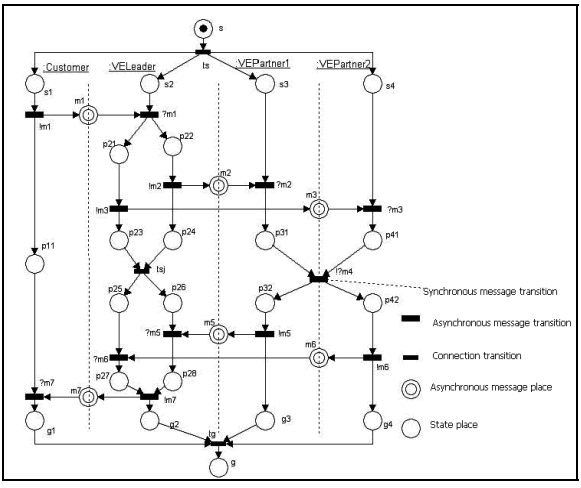


Fig. 2. The interactive Petri net model of VE-GBP

3 Modeling of the Sub-business Process of Member Enterprises (ME-SBP)

3.1 Establish a Model of Detailed Sub-business Process of Member Enterprises

Under the constraints of the interactive interface model of sub-business process of member enterprises, each member enterprise has its own detailed sub-business

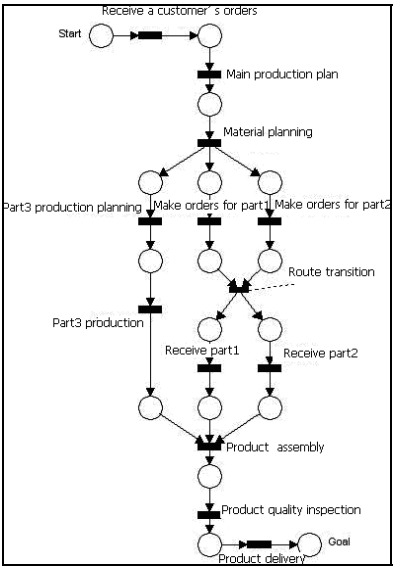


Fig. 3. The Petri net model of VE-SBP

processes. Petri net can be used to effectively describe the various activities in their sub-business processes and their executing sequence.

In order to clearly describe the start and end of sub-business process, a starting inventory Star and a target inventory Goal are introduced. In addition, when Petri net is used to describe the sub-business process, in order to realize the full description of the various logic relations among activities, route transition should also be introduced and the route transition does not correspond to any concrete activity. Figure 3 shows the Petri net model of the sub-business process of the core enterprise VELeader.

3.2 Establish Object-oriented (O-o) Petri Net Model of ME-SBP

In order to meet the confidential requirements of member enterprises for their own business processes, object-oriented Petri net model of ME-SBP is used to carry out encapsulation on the model of sub-business process. First, on the basis of the interactive model of VE-BP, all input/output messages involved in the interactions of the member enterprises with the outside world can be obtained and each input/output message is represented as an input/output message place. Second, each input/output message of the member enterprises certainly has direct reception/sending relation with certain activities in the model of sub-business process and in this paper such activities are called input/output interface activities of sub-business process. Then, the input/output message places and their corresponding transitions of input/output interface activities are connected. Through such treatment, input/output message places and their directly related transitions of input/output interface activities constitute the external interface of the sub-business process model while the other parts of the Petri net model of sub-business process constitute the internal model of the member enterprises, which is no seen by the outside world. In order to distinguish the internal model and external interface of the sub-business process and realize the correct connection of the two, object-oriented Petri net model of ME-SBP can be established. The internal model deals with the members' private messages which can be seen only by the member enterprises, while the external interface is constituted by input/output messages and input/output interface activities that can be seen by others besides the member enterprises, which is the medium of communication between member enterprises and other enterprises. Through the object-oriented Petri net model, not only the encapsulation of the model of ME-SBP is realized but also the message communication between member enterprises and other enterprises is described in all aspects.

Now a formal definition will be made on object-oriented Petri net model of ME-SBP, to describe all the component elements of the model. Object-oriented Petri net model of ME-SBP is a sextuple.

$OOPN_i = (PS_i, PM_i, TS_i, TM_i, TR_i, F_i)$, in which:

(1) PS_i represents the collection of state places;
(2) PM_i represents the collection of all input/output message places;
(3) TS_i represents the collection of all the transitions of internal activities in sub-business processes;

(4) TM_i represents the collection of transitions of all the input/output interface activities in the sub-business processes;
(5) TR_i represents the collection of all the route transitions in the sub-operation process model;
(6) F_i represents the flow relation between places and transitions.

The object-oriented technology is applied and the six components are represented by the concept "class" in the objective-oriented language C++. The hierarchy structure of classes is shown in Figure 4.

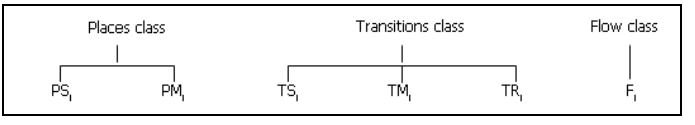


Fig. 4. The hierarchy structure of classes

Then, place class
Attributes: token-name
Message place class
Attributes: station-no
station-state
Operation: get-station-state()
set-station-state()

Other classes can be defined in the same way. Through the application of the object-oriented technique, the internal business process of each member enterprise can be encapsulated to be invisible to the outside world, but the interface activities and

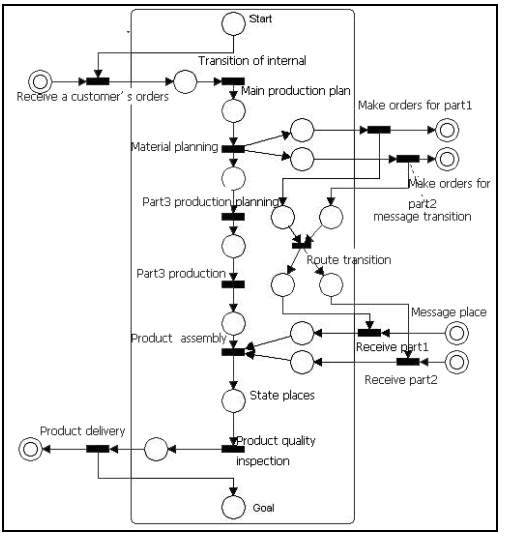


Fig. 5. The o-o Petri net model of VE-SBP

message exchange among members are open to the public. At the same time, the application of this technique can also enhance the maintainability and reusability of the system model. In the following Figure 5, the object-oriented Petri net model of VELeader-SBP is shown. The internal business process in the square frame is encapsulated while the process outside of the square frame is open.

According to this way, models of the sub-business process of other member enterprises can be established. If these models are connected on the basis of the constraining requirements of the interactive model, then the object-oriented Petri net model of whole business process of the supply of virtual enterprises can be established. Because there are a lot of member enterprises in the supply chain of a virtual enterprise, model for each member enterprise will not be introduced. This paper would like to illustrate this modeling method with above example.

4 Emulation Analysis on the Reasonability of the Model

Through above method, model of VE-GBP and model of ME-GBP are established. Then emulation qualitative analysis can be conducted on the established Petri net model. Major problems in the qualitative analysis are whether the system is locked and whether the process ends in a correct way. According to reference [1], a business process model should meet the following two conditions to be considered as reasonable:

(1) the business process can realize the whole process from start to end;
(2) No dead locking or endless loop will appear in the execution of the business process.

From above two points, it can be seen that emulation method can be used for analyzing the reasonability of a business process model.

As for the model of sub-business process, the initial state of the system is that: only the staring place Start has a token. The normal ending state of the system should be: only the targeting place Goal has a token. In order to verify the reasonability of a model, emulation running is the simplest method. This paper carries emulation on the

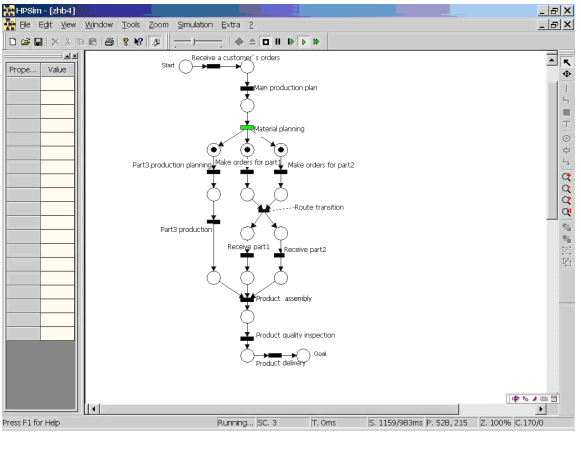


Fig. 6. Petri net emulation interface

Petri net model of VELeader-SBP in Figure 3 and the test-version Petri net emulation software Hpsim is used. The emulation interface is shown in Figure 6. Results of the emulation illustrate that the model meets above two conditions, so the model of VELeader-SBP is partially reasonable.

The reasonability of the interactive model of business process in Figure 2 can also be verified in the same way. The initial state of the system is: only the staring place Start has a token. The normal ending state of the system should be: only the targeting place Goal has a token. The emulation interface is shown in Figure 7. Results of the emulation illustrate that the interactive model also meets above two conditions, so the interactive model of VE is also reasonable.

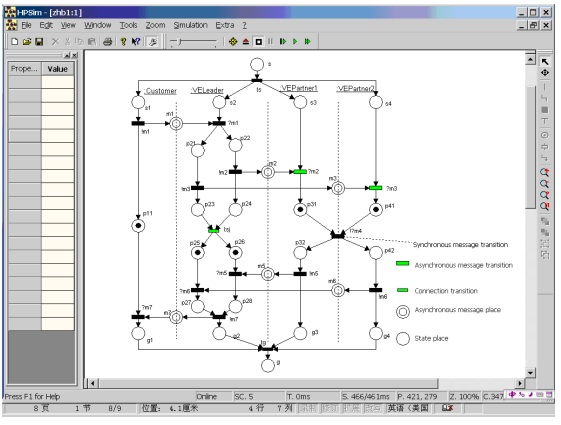


Fig. 7. Petri net emulation interface

5 Conclusion

In this paper, UML and object-oriented Petri net are combined to establish the model for the distributed business process of the supply chain of virtual enterprises. On the one hand, the confidentiality of enterprises are considered and on the other hand the problems of models established by using ordinary Petri net, such as large size and state explosion of system, are solved. Beyond that, Petri net emulation tool Hpsim is applied to carry out emulation running, to verify whether the business processes of the supply chain of virtual enterprises are smooth or not, whether the established models are reasonable or not and whether there is any bottleneck in existing business processes.

References

1. van der Aalst, W., van Hee, K.: Work Flow Management-Model, Method and System, pp. 231–239. Tsinghua University Press, Beijing (2004)
2. Chen, B.B.: Supply Chain Management: Strategy, Technique and Practise, pp. 15–50. Electronics Industry Press, Beijing (2004)

3. Yao, M.J., Chiou, C.C.: On a replenishment coordination model in an integrated supply chain with one vendor and multiple buyers. European Journal of Operational Research 159(2), 406–419 (2003)
4. Axsater: Approximate optimization of a two-level distribution inventory system. Int. J. Production Economics 12, 37–40 (200)
5. Yu, J., Shen, K., Huang, B.: Application of Coloured Petri Net in Modeling of Work Flow. Computer Applications and Software 10(7), 381–384 (2004)
6. Ma, X.Y., Jiang, C.J.: Design and Analysis of SCM Using TPN. Computer Engineering 29(1), 82–84 (2003)
7. Jain, A.: Performance modeling of FMS with flexible process plans-a Petri net approach. International Journal of Simulation Modelling 5, 101–113 (2006)
8. Liu, C., Kondratyev, A., Watanabe, Y., Desel, J., Sangiovanni-Vincentelli: Schedulability analysis of Petri nets based on structural properties. Applications of Concurrency to System Design (ACSD), 69–78 (2006)
9. Kabicher, S., Kriglstein, S., Rinderle-Ma, S.: Visual Change Tracking for Business Process Models. In: Jeusfeld, M., Delcambre, L., Ling, T.-W. (eds.) ER 2011. LNCS, vol. 6998, pp. 504–513. Springer, Heidelberg (2011)
10. Datta, S.: Modelling TMCP of HSLA steels by Petri neural network technique. Canadian Metallurgical Quarterly 45(3), 303–310 (2006)
11. Smirnov, S., Weidlich, M., Mendling, J., Weske, M.: Action Patterns in Business Process Models. In: Baresi, L., Chi, C.-H., Suzuki, J. (eds.) ICSOC-ServiceWave 2009. LNCS, vol. 5900, pp. 115–129. Springer, Heidelberg (2009)
12. List, G.F., Cetin, M.: Modeling traffic signal control using Petri nets. IEEE Trans. on Intelligent Transportation Systems 5(3), 177–187 (2004)
13. Velardo, F.R.: Coding Mobile Synchronizing Petri Nets into Rewriting Logic. Electronic Notes in Theoretical Computer Science 4, 83–98 (2007)

Tree Branching Reconstruction from Unilateral Point Clouds

Yinghui Wang[1], Xin Chang[1], Xiaojuan Ning[1], Jiulong Zhang[1], Zhenghao Shi[1], Minghua Zhao[1], and Qiongfang Wang[2]

[1] Institute of Computer Science and Engineering
Xi'an University of Technology, Xi'an, China, 710048
[2] Faculty of Automation and Information Engineering
Xi'an University of Technology, Xi'an, China, 710048
{wyh,changxin,ningxiaojuan,zhangjiulong,
ylshi,zhaominghua,qfwang}@xaut.edu.cn

Abstract. Trees are ubiquitous in natural environment and realistic models of tree are also indispensable in computer graphics and virtual reality domains. However, their complexity in geometry and topology make it a great challenge for photo-realistic tree reconstruction. Since tree trunk is the preliminary structure of trees, its modeling is a critical step which plays an important role in tree modeling. Many existing methods focus on the overall resemblance of tree branches but omit the local geometry details. In this paper, we perform unilateral scanning of real-world trees and propose an approach that could reconstruct trees from incomplete point clouds. The core of our method contains four parts: local optimal segmentation of tree branch, skeletal point and lines extraction from unilateral branch, the cross-section construction of tree branch, and final tree branch surface generation. Experimental results demonstrate the effectiveness and robustness of our method which could keep realistic shape of trees.

Keywords: Unilateral point cloud, Branch Modeling, Skeleton Line, Cross-section construction.

1 Introduction

Generating accurate tree models could play an essential role in virtual reality, animated cartoon, city planning and many other fields. However, intricate instances of branching and organ structures with varied postures and spatial distributions, coupled with the difficulty in acquiring the geometric information from natural environment, make it a much more challenging work.

Works on tree modeling can be regarded as reconstructing the tree trunk and then synthesizing tree leaves. For effective modeling, dominant approaches are proposed which could be summarized into the following categories: cylindrical fitting [4],[23], adjacency graph [34], mathematical morphological method [9], volume decomposition [16], Octree division method [5], Octree-Graph [1], Level sets [32] and etc.

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 250–263, 2012.
© Springer-Verlag Berlin Heidelberg 2012

In this paper, we are interested in solving two problems that appeared in tree modeling. Firstly, most methods concentrate on the overall fidelity of tree shapes but neglect the local accuracy of tree branch. In addition, recent work mainly utilized the data registered by multiple viewpoints scanning which leads to costly computation. Therefore, we proposed a method that suitable for unilateral scanned tree data and could achieve robust skeleton model of trees.

Our method naturally lends itself to efficient tree trunk reconstruction of unilateral scanned data, which is efficient to construct a more accurate and detailed hierarchical tree skeleton. We demonstrate that our method succeeds in generating tree model by local optimal segmentation, skeleton abstraction and connection, and surface model generation.

2 Related Work

By far, the methods for tree modeling can be mainly classified into two categories: virtual trees designing and real tree modeling.

Virtual Trees Designing
The main idea of designing virtual trees is to simulate the growth of trees, and many typical methods are based on imaginary trees.

Characteristic methods such as AMAP model [25], L-system [21,27], Green Lab model [20], procedural modeling [29,2] and so on. These methods are mainly based on some biological model and geometric rules [25], L-system [21] , and some provided human interaction interface [20] or scratch-based branch models [3,10,19]. Although these methods can produce more vivid 3D tree models and can be applied in the virtual scene [22], most of the trees created are imaginary. As a result, the 3D models of the real tree cannot be generated.

Real Trees Modeling
A number of reconstruction methods have been developed that allow for modeling trees from real world data such as sets of images or scanned data.

Image-based tree reconstruction mainly recovers the 3D information from a series of images taken at multiple angles [8,31,17]. Much more effective methods are proposed, for example, billboard-type modeling [28], middle axis estimation [26], volumetric-based modeling [24], particle simulation [18], model based matching method [6,30] and physical based method [35] . However, the high inaccuracy and difference between the model and given trees make the modeling an insurmountable problem. Besides, since multiple angle data acquisition is not always successful due to occlusion in trees, new acquisition and reconstruction methods of trees must be explored.

3D laser scanner is a new digital capturing device after digital cameras. This enables effective modeling of complex shaped objects such as trees, as well as prepares the data needed for developing the complex shape modeling theory [15]. Xu et al. [33] proposed to cluster the edges in spanning graph to reconstruct the tree skeleton. Gorte [9] considered the skeletonization of trees in production orchards. Cheng et al. [7] reconstructed the tree branches and crown shapes based

on range image. Cote et al. [32] generated tree based on scattering properties obtained from different intensities of point. Livny et al. [13] utilized global optimizations to reconstruct the branching structure of multiple overlapping trees. Livny et al. [12] presented a lobe-based tree representation for modeling trees, in which a supervised classification method is developed to classify trees.

3 Method

3.1 Local Optimal Segmentation of Tree Trunk

We utilize a filtering algorithm [11] to acquire a local optimal segmentation method of tree branches. The method is a simple and efficient implementation of Lloyd clustering algorithm [14] based on k-d tree.

Local optimal segmentation method has advantages over random segmentation shown in Figure 1, which could provide more intuitive and suitable segmentation results. The algorithm begins by storing the data points in k-d tree. Each node of k-d tree is associated with a closed box, called cell, and then execute the following steps:

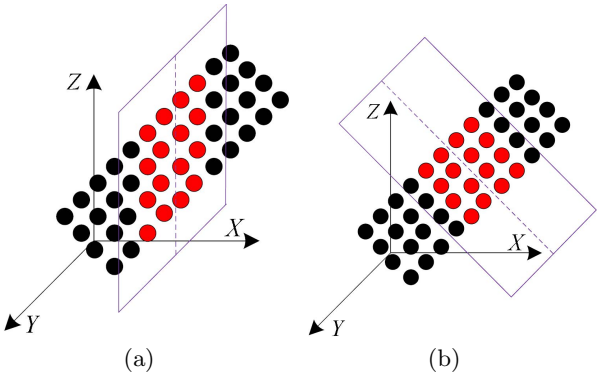


Fig. 1. Segmentation Comparison. (a) random segmentation, (b) local optimal segmentation.

(1) Compute the candidate $z^* \in Z$ that is closest to the midpoint of C, where C represents the cell and Z is the candidate set.

(2) For each of the remaining candidates $z^* \in Z \setminus \{z^*\}$, if no part of C is closer to z^*, we can infer that z is not the nearest center to any data point associated with u and hence we can filter z from the list of candidates.

(3) If u is associated with a single candidate (which must be z^*) then z^* is the nearest neighbor of all its data points. We can assign them to z^* by adding the associated weighted centroid and counts to z^*.

(4) If u is an internal node, we recurse on its children.

(5) If u is a leaf node, we compute the distances from its associated data point to all the candidates in Z and assign the data point to its nearest center.

3.2 Tree Skeleton

The skeleton of tree is a free curve by orderly connecting skeleton points, which can not only depict the distorted form of the branches, but also coincides with the centralized axis of the branches.

Skeleton Definition
Different trees may have complicated shapes. Skeleton can be described based on each segmentation surface by three basic geometrical parameters (see Figure 2): central point p, optimal axis direction α and the radius r, denoted as $S(p, \alpha, r)$. We treat the centroid of optimal segmentation as skeleton point which satisfies the fact that the skeleton line passes through the center of branches. In fact, it must coincide with the following principles.

Position. Skeleton point $p(x, y, z)$ located on the local cross-section plane of branch and the point q_i belongs to the local surface point set. They may satisfy the condition that p and q_i may have minimum distance towards the direction of normal vector shown in Eq. (1).

$$p(x, y, z) = \arg\min \sum_{p_i \in N} \|(p - q_i) \times n(q_i)\| \tag{1}$$

Direction. The axis direction of skeleton is the minimum sum of angles with the normal of boundary points on cross-section, as shown in Eq. (2).

$$\alpha = \arg\min \sum_{p_i \in N} \cos^{-1}(\alpha \cdot n(q_i)) \tag{2}$$

Radius. The radius of branch is the minimum distance from p to boundary points q_i, defined by Eq. (3).

$$r = \min\{\|p - q_i\|\}\, q_i \in \Pi_{q_i} \tag{3}$$

Where $n(q_i)$ denotes the normal vector of point q_i, and Π_{q_i} is the point set.

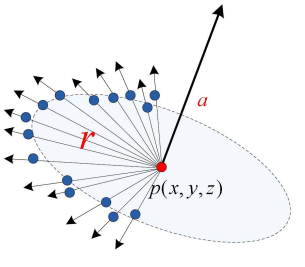


Fig. 2. Skeleton Parameters

3.3 Skeleton Points Extraction

In this paper, we extract the skeleton points by circle fitting method. As described above, skeleton points located on the local cross-section plane H of tree branch which could be approximated as a series of circular (by least squares fitting method). H can be determined by the centroid point μ and the normal $\bar{n}$ of each segment, where $\mu = 1/N \sum_{i=1}^{N} q_i$, and $\bar{n}$ is calculated by the maximum eigenvectors of the covariance matrix $M = 1/N \sum_{k=1}^{n} (q_k - \mu)(q_k - \mu)^T$. We assume eigenvalues $\lambda_0 \leqslant \lambda_1 \leqslant \lambda_2$, the corresponding eigenvectors e_0 and e_1 could define H.

We project the each segmentation point set into respective H and then approximate these projection points as a circle. The fitting circle can be achieved by the following steps. Therefore, the skeleton points could be extracted from the center of the circle.

(1) The normal vector of H is rotated around Z axis, and parallel with plane $X - Y$;

(2) Project each of the segmentation set onto the plane $X - Y$;

(3) Then we can obtain the circle by least square fitting on the plane $X - Y$, i.e. the fitting must satisfy:

$$d(x_i, y_i) = \sqrt{(x_i - x)^2 + (y_i - y)^2} - r \tag{4}$$

(4) Based on the circle in 2D, we restore the circle center point and the radius in 3D by coordinate transformation. As in Eq. (5), $\{x_c, y_c, z_c\}$ denotes the position of plane H.

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ x_c & y_c & z_c & 1 \end{bmatrix} \tag{5}$$

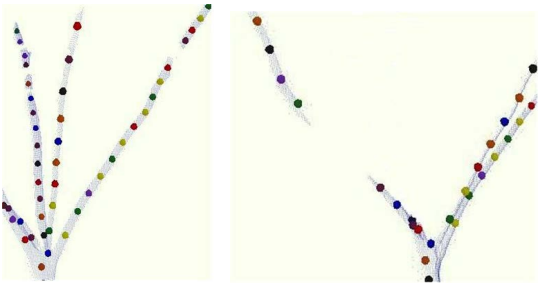
(5) Rotate the normal of the circle that consistent with that of plane H.

3.4 Skeleton Lines

There are inevitably occlusions during data collection thus leading to the discontinuity of branch in skeleton. Therefore, it is an essential problem to connect the broken skeleton lines. Discontinuous branches may have a gap between skeleton points, shown in Figure 3. So the most important is to find the parental skeleton points located before bifurcation. On the contrary, it is effortless to implement the skeleton line of continuous branch.

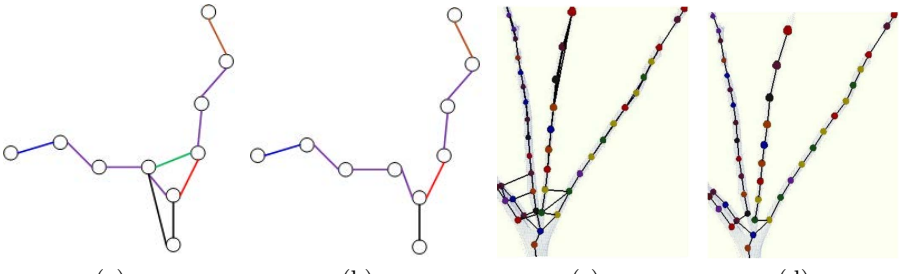
Skeleton Line of Continuous Branch

For continuous branch, we could reconstruct tree skeleton by the adjacency graph between segments (Figure 4(a)). The adjacency graph is generated by the neighboring points for each segment. Each point is compared with its neighbors, and



(a) (b)

Fig. 3. Skeleton line



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

Fig. 4. Skeleton line. (a) Adjacency graph, (b) minimum spanning tree, (c) skeleton using (a), (d) skeleton using (b).

if they belong to different segments then we consider that they are neighbors, and skeleton calculated from adjacency graph is illustrated in (Figure 4(c)). Due to the condition that the adjacency graph may have loops, thus it is extended to minimum spanning tree which is an effective way to remove loop and connects the graph nodes with minimum weight in the loop shown in Figure 4(b). Here the weight is taken as Euclidean distance between points. Figure 4(d) is the minimum spanning tree of the skeleton adjacent graph, which has eliminated the graph loop completely.

Skeleton Line of Discontinuous Branches

The skeleton line of discontinuous branches is achieved by searching its parent point the main skeleton line is to find which must be adjacent and at near end of the main skeleton, meanwhile their axis is also similar. For instances, for the first disconnected skeleton point B in Figure 5, if there exists a neighboring skeleton point A lying on the main skeleton and their axis angle is less a threshold (in the following the threshold is 10°), then we can connect point A and B (as shown in red) and thus the separated points are connected to the main branch. Figure 5 shows the connection result implemented by the similarity in skeleton axis for

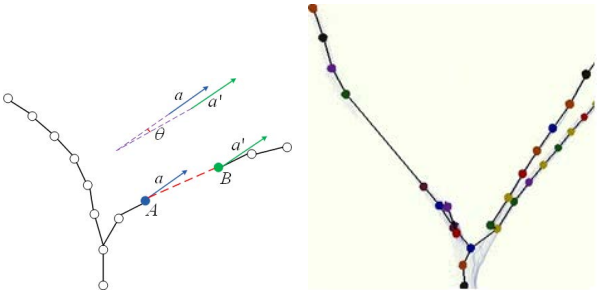


Fig. 5. Skeleton of Discontinuous Branches. (a) Angles consistency, (b) Skeleton line.

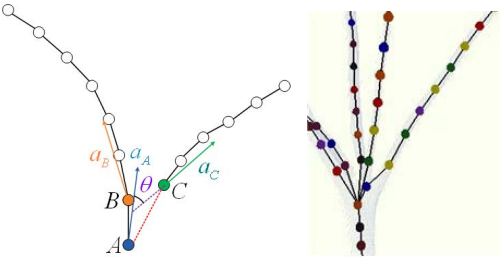


Fig. 6. Skeleton Lines at Bifurcation. (a) Bifurcation angles, (b) Skeleton Lines.

Figure 3(b). We can see that the resulting axis has the same trend with the original branch.

Skeleton Line at Bifurcation

At bifurcation, the distances between the children branches are usually less than that between the parent branches, which thus leading to wrong connection when using minimum spanning tree in Figure 4(b). To tackle this problem, we adopt the upward searching method to find the true parent skeleton point for separated point at bifurcation.

Generally, the angle between the axis of the parent branch and the child branch is less than 45° for all the bifurcation branches, which is used to connect skeleton axis at bifurcation. As can be seen from Figure 6(a), the children skeleton point C at bifurcation has a brother point B, and the distance between C and B is less than that between C and A. Thus B is chosen as the parent point, and obviously the angle between α_C and α_B is larger than 45°. Here the upward search method is used to find parent point, i.e., start searching from the separated point B. If the angle between wrong parents points B and its parent point A is less than , then we consider point A as its real parent. So it is reasonable to disconnect B from C and connect A to C (red line). Otherwise, continue this

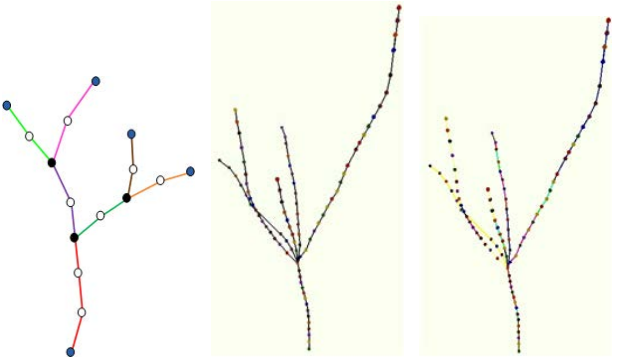


Fig. 7. Skeletons Segmentation. (a) Rules, (b) and (c) show before and after segmentation.

search until proper parent is found. Figure 6(b) shows the result of removing wrong connections in Figure 4(d), which better maintains trees structure.

4 Tree Branch Reconstruction

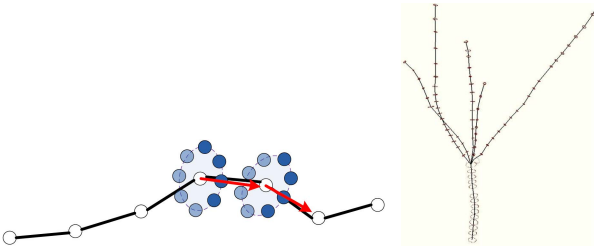
4.1 Skeleton Segmentation between Different Branches

Whatever complex the branch morphology is, the bifurcation skeleton can be used to separate different branches. Therefore, the root point does not have parents, bifurcation point may have more than two adjacent points and the ending skeleton point does not have children. According to these principles, it is convenient to detect the bifurcation points which could provide an efficient way to detect different branches and make it easier to construct tree branch model. Figure 7(a) shows segmentation rules with bifurcation points. Figure 7(b) is the original branch skeleton, and Figure 7(c) is the separation result where different colors represent different segments of the branch skeletons.

4.2 Tree Branch Model

To acquire the complete branch, we re-sample the skeleton cross-section to substitute the true points. The re-sampling is achieved by counter clock wisely along the cross-section centering on the skeleton point axis. Three geometric parameters of skeleton point are used to re-sample in order to generate points on the local cross-section surface.

Figure 8(a) describes the re-sampling rule and Figure 8(b) displays the re-sampling results. The circle that sampling points located on could illustrate the thickness of tree branches. Moreover, these circles are perpendicular to skeleton lines which are helpful for lifelike tree models.



(a) (b)

Fig. 8. Re-sampling process. (a) The re-sampling rules of local cross-section, (b) the instances of re-sampling results.

Obviously, the triangle surface is composed of points on two adjacent cross-sections. A valid triangle surface should meet the following three conditions:

(1) Three vertices of each triangle cannot be from the same cross-section;
(2) The next vertex is chosen according to the triangle that has an outer circle with minimum radius;
(3) Two adjacent triangles form a convex quadrilateral.

Then these vertices are connected according to the aforementioned rules, we can obtain the mesh model of tree branch.

5 Experimental Results and Analysis

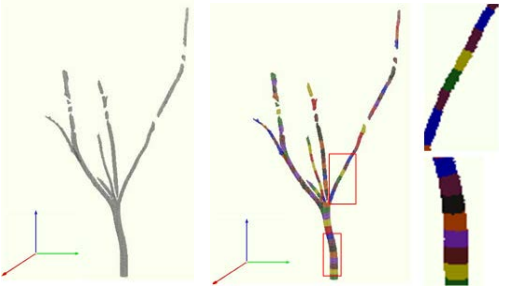
All experiments were performed on a PC with Intel Pentium D processor and 2GB memory.

5.1 Branch Reconstruction

In this section, we demonstrate the efficiency of our branch modeling method. Figure 9 displays the result of the local optimal segmentation of tree trunk. We can conclude that our method could make better segmentation according to growth direction of tree. It is irrelevant with tree shape.

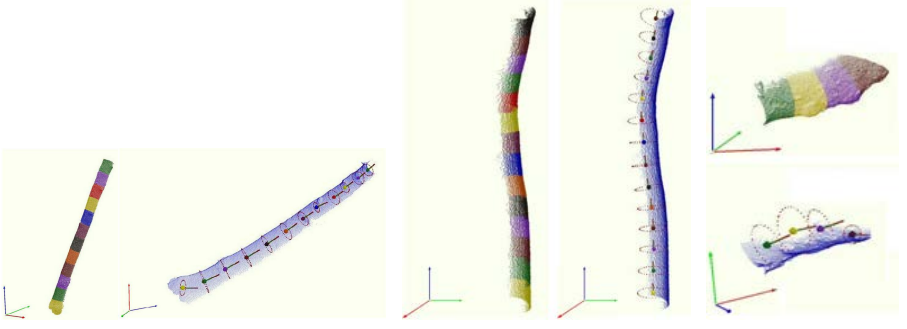
Figure 10 is an instance using the proposed method described herein to extract skeletal point from the branches in different direction. Different colors indicate different optimal segmentation point set, and points with corresponding color indicate the location of skeletal point. The lines attached to the skeleton points are their axis. The circumference consisted by red dot which means the circular cross-section drawn from the radius of skeleton points. Experimental results indicate that our method could deal with trees with intricate structure.

Figure 11 shows the whole steps of branch modeling from five perspective views. Seen vertically, there are original cloud data, local optimal trunk segmentation, skeleton extraction and connection, cross-section of branches, and the triangular mesh model (see in Figure 12).



(a) (b)

Fig. 9. Local optimal segmentation. (a) Tree branch, (b) branch segmentation.



(a) (b) (c)

Fig. 10. Skeleton Extraction. (a) Inclined branch, (b) vertical branch, (c) horizontal branch.

5.2 Error Analysis

In general, the branches expressed by single-scanned cloud data are in three forms: straight, broken and bent. For straight branch, the approximation result is usually good; but for broken and bent branch, the approximation may have noticeable error. To evaluate the errors, we use two additional data that the circumferences at the beginning and ending point of skeleton line besides that at the broken, before and after bent points.

Figure 13 is a comparison between the summed distance of the original scan and the reconstructed branch at the above selected five locations. Since only part of the branch is obtained, we choose 2/5 or 3/5 the circumference of cross-section according to the original scan. In Figure 13, the red line stands for the reconstructed result, the blue line for the original data. Although there is some error due to noise or other factors, the average error is 3.3mm, which infers that it has a good fidelity with the real data.

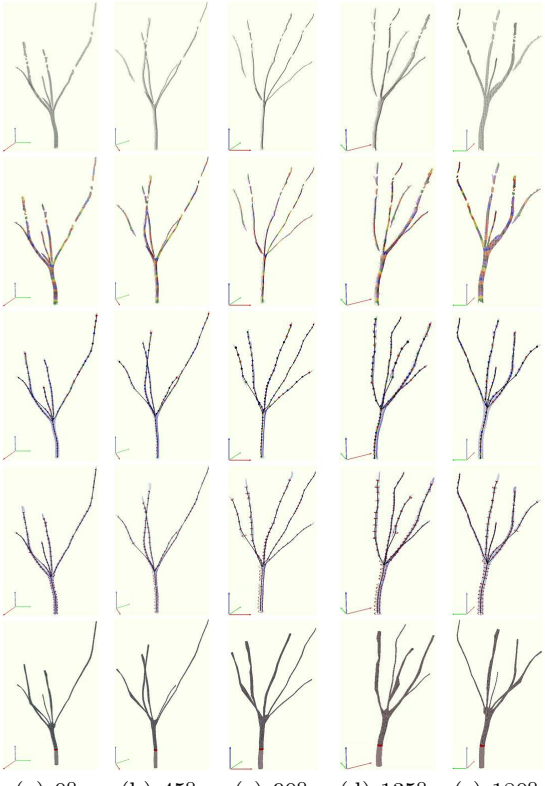


Fig. 11. Different Perspectives of Branch Models

Fig. 12. A bark model described by triangular mesh

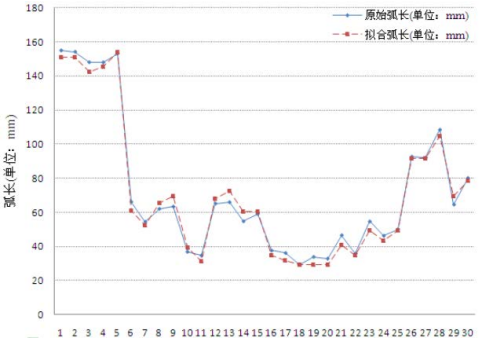


Fig. 13. Original arc length and the fitting arc length analysis

6 Conclusions

In this paper, we proposed a method of trees skeleton modeling which is more accurate to represent the skeleton point for unilateral scanned data. Our method may have several advantages, firstly it could deal with the point data directly without classifying all the tree branches, and it just treats the tree branches integrally; secondly, the geometric parameters of the skeleton points are closer to original tree branch; thirdly, it is more effective to solve the condition when the branches are occluded; furthermore it is efficient and accurate to regulate the skeletal lines especially the Bifurcate skeleton. Experimental results demonstrate the effectiveness and efficiency of our proposed method, which could provide lifelike tree models in the nature.

Acknowledgment. This work is supported in part by National Natural Science Foundation of China under Grant No.61072151; and in part by Shaanxi Educational Scientific Research Plan under Grant No.09JK628, 2010JK734 and 2010JK700; and in part by Shaanxi Scientific Research Plan under Grant No 2011K06-35.

References

1. Alexander, B., Roderik, L.: Campino-a skeletonisaztion method for point cloud preprocess 63, 115–127 (2008)
2. Bedřich, B., Ondřej, Š., Radomír, M., Gavin, M.: Guided Procedural Modeling. In: Eurographics, vol. 30, pp. 325–334 (2011)
3. Blaise, F., Barczi, J.F., Jaeger, M., Dinouard, P., De Reffye, P.: Simulation of the growth of plants. Modeling of metamorphosis and spatial interactions in the architecture and development of plants, pp. 81–109. Springer (1998)
4. Bloomenthal, J.: Modeling the mighty maple. SIGGRAPH Comput. Graph. 19, 305–311 (1985)

5. Bucksch, A., Lindenbergh, R., Menenti, M.: Skeltre-fast skeletonization of imperfect point clouds of botanic trees. In: Eurographics/ACM Siggraph Symposium on 3D Object Retrieval, pp. 13–20 (2009)
6. Chen, X., Neubert, B., Xu, Y.Q., Deussen, O., Kang, S.B.: Sketch-based tree modeling using markov random field. ACM Trans. Graph. 27, 109:1–109:9 (2008)
7. Cheng, Z.L., Zhang, X.P., Chen, B.Q.: Simple reconstruction of tree branches from a single range image. Journal of Computer Science and Technology 22, 846–858 (2007)
8. Forsyth, A,D., Pronce, J.: Computer Vision: a modern approach, vol. 1. Prentice Hall (2002)
9. Gorte, B.: Skeletonization of laser-scanned trees in the 3d raster domain. In: Innovations in 3D Geo Information Systems. Lecture Notes in Geoinformation and Cartography, pp. 371–380. Springer, Heidelberg (2006)
10. Ijiri, T., Owada, S., Igarashi, T.: The Sketch L-System: Global Control of Tree Modeling Using Free-Form Strokes, vol. 1, pp. 138–146. Springer (2006)
11. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: An efficient k-means clustering algorithm: Analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell. 24, 881–892 (2002)
12. Livny, Y., Pirk, S., Cheng, Z., Yan, F., Deussen, O., Cohen-Or, D., Chen, B.: Texture-lobes for tree modelling. ACM Trans. Graph. 30, 53:1–53:10 (2011)
13. Livny, Y., Yan, F., Olson, M., Chen, B., Zhang, H., El-Sana, J.: Automatic reconstruction of tree skeletal structures from point clouds. In: ACM SIGGRAPH Asia 2010 papers, SIGGRAPH ASIA 2010, pp. 151:1–151:8. ACM (2010)
14. Lloyd, S.: Least squares quantization in pcm. IEEE Transactions on Information Theory 28(2), 129–137 (1982)
15. Long, C., Zhao, J., Goonetilleke, R.S., Xiong, S., Ding, Y., Yuan, Z.Y., Zhang, Y.: A new region growing algorithm for triangular mesh recovery from scattered 3d points. T. Edutainment 6, 237–246 (2011)
16. Ma, W., Xiang, B., Zhang, X., Zha, H.: Decomposition of branching volume data by tip detection. In: Proceedings of IEEE International Conference on Image processing, pp. 1845–1851 (2008)
17. Maxime, L., Long, Q.: A quasi-dense approach to surface reconstruction from uncalibrated images. IEEE Trans. Pattern Anal. Mach. Intell. 27, 418–433 (2005)
18. Neubert, B., Franken, T., Deussen, O.: Approximate image-based tree-modeling using particle flows. ACM Trans. Graph. 26 (July 2007)
19. Okabe, M., Owada, S., Igarashi, T.: Interactive design of botanical trees using freehand sketches and example-based editing. Computer Graphics Forum 24, 313–318 (2005)
20. Oliver, D., Bernd, L.: A modeling method and user interface for creating plants. Graphics Interface 17, 189–197 (1997)
21. Prusinkiewicz, P., Aristid, L.: The Algorithmic Beauty of Plants. Springer, New York (1990)
22. Pan, J., Zhang, J.J.: Sketch-based skeleton-driven 2d animation and motion capture. Transaction on Edutainment 6, 164–181 (2011)
23. Pfeifer, N., Gorte, B., Winterhalder, D.: Automatic reconstruction of single trees from terrestrial laser scanner data. In: Proc.ISPRS Conf., Int. Archives of Photogrammetry and Remote Sensing, Instanbul, Turkey, vol. 35, pp. 114–119 (2004)
24. Reche-Martinez, A., Martin, I., Drettakis, G.: Volumetric reconstruction and interactive rendering of trees from photographs. ACM Trans. Graph. 23, 720–727 (2004)

25. de Reffye, P., Edelin, C., Françon, J., Jaeger, M., Puech, C.: Plant models faithful to botanical structure and development. SIGGRAPH Comput. Graph. 22, 151–158 (1988)
26. Shlyakhter, I., Rozenoer, M., Dorsey, J., Teller, S.: Reconstructing 3d tree models from instrumented photographs. IEEE Comput. Graph. Appl. 21, 53–61 (2001)
27. Stava, O., Benes, B., Mech, R., Aliaga, D.G., Kristof, P.: Inverse procedural modeling by automatic generation of l-systems. Graph. Forum 29, 665–674 (2010)
28. Stephen, E., Hope, H.: Landscape Modeling: Digital technology for landscape visualization. McGraw-Hill Professional Publishing (2001)
29. Talton, J.O., Lou, Y., Lesser, S., Duke, J., Měch, R., Koltun, V.: Metropolis procedural modeling. ACM Trans. Graphics 30 (2011)
30. Tan, P., Zeng, G., Wang, J., Kang, S.B., Quan, L.: Image-based tree modeling. ACM Trans. Graph. 26 (July 2007)
31. Teng, C.H., Chen, Y.S., Hsu, W.H.: Constructing a 3d trunk model from two images. Graph. Models 69, 33–56 (2007)
32. Cotø, J.-F., Widlowski, J.-L., Fournier, R.A., Verstraete, M.M.: The structural and radiative consistency of three-dimensional tree reconstructions from terrestrial lidar. Remote Sensing of Environment 113, 1067–1081 (2009)
33. Xu, H., Gossett, N., Chen, B.: Knowledge and heuristic-based modeling of laser-scanned trees. ACM Trans. Graph. 26 (October 2007)
34. Yan, D., Wintz, J., Mourrain, B., Wang, W., Boudon, F., Godin, C.: Efficient and robust reconstruction of botanical branching structure from laser scanned points. In: 11th IEEE International Conference on Computer-Aided Design and Computer Graphics (CAD/Graphics), pp. 572–575 (2009)
35. Yang, M., Huang, M.-c., Wu, E.-h.: Physically-Based Tree Animation and Leaf Deformation Using CUDA in Real-Time. In: Pan, Z., Cheok, A.D., Müller, W. (eds.) Transactions on Edutainment VI. LNCS, vol. 6758, pp. 27–39. Springer, Heidelberg (2011)

Literature Analysis on the Higher Education Informatization of China (2001-2010)*

Qiaoyun Chen and Xiaoyan Qiao

School of Educational Sciences, Nanjing Normal University
No.122, Ninghai Road, Nanjing, Jiangsu, China, 210097
qiao7@163.com

Abstract. This paper selects 278 journal articles from CNKI between 2001 and 2010 on higher education informatization. A statistical result is given from four dimensions: basic theory, construction, management and evaluation of higher education informatization. The authors analyze the statistical result from the distribution of article amount, authors and contents of the 278 articles.

Keywords: university informatization, higher education, Content analysis.

1 Introduction

Higher education informatization is a huge project. From the macro perspective, it relates to the management of higher education institutions, teaching, research and social services, etc.; from the micro, it includes the construction of higher education institutions, information infrastructure, teaching resources, personnel, management system and so on. Since the 1990s, all countries in the world are accelerating the pace of higher education informatization. China has taken a series of important measures to develop informatization in education, and higher education scholars carry out relevant researches from different aspects.

2 Study Method

In this paper, "higher education informatization", "university informatization" and "college informatization" are used as the keywords. We use the exact match search method for journal articles between 2001 and 2010 in the CNKI (China National Knowledge Infrastructure). There are 278 papers in total excluding news, interview, monographs network education, and records management. We analyze these papers in 4 dimensions: basic theory, informatization construction, and informatization

* This research was supported by the National Education Sciences "12th Five-Year" Planning Ministry of Education youth issues, *Research on use benefit of Higher education informatization based on user satisfaction* (No.ECA110332).

Z. Pan et al. (Eds.): Transactions on Edutainment VIII, LNCS 7220, pp. 264–273, 2012.
© Springer-Verlag Berlin Heidelberg 2012

management and informatization evaluation. The statistical tool is Microsoft Office Excel 2007. We can see the results listed in table 1.

Table 1. Research Content of Higher Education Informatization

	aspect	Number of papers	Sum
Basic Theory	overview on University informatization	16	115
	current status, trends and suggestions on university informatization development	35	
	university informatization construction system/mechanism	4	
	problems and countermeasures	24	
	education reform under the university informatization	10	
	impact of University informatization on teachers	5	
	foreign experiences study	21	
Informatization Construction	infrastructure of university informatization construction	8	98
	Application System	73	
	Information Resources	7	
	Organizational team building	5	
	Information standard building	2	
	Information Services	3	
Informatization Management		26	26
Informatization Evaluation		39	39

3 Result and Discussion

3.1 Analysis of Number Distribution

Generally speaking, the number of papers related to higher education informatization between 2001 and 2010 distributes like a ladder (see Figure 1). The number of papers on informatization management and information technology evaluation does not change much during the ten years, which shows that the main characteristics of the ladder is two dimensions of the basic theory and informatization construction. From the curves in Figure 2, we can learn: first, the research of higher education informatization shows a increasing trend; second, related study presents a polarization state. The basic theory and informatization construction get the attention of the academic research, and the number of published articles increased yearly. On the contrary, informatization management and evaluation have been in a stable status, which forms the two extremes with the other two aspects. In 1999, the Ministry of Education promulgated the "21st Century Education Revitalization Action Plan", and China set off a magnificent wave of information technology education, and has made substantial breakthrough in infrastructure, resource development, key technologies and major applications. Through 10 years of development, the research has been mature, and the theory and practice are always complementary, so the number of

papers on the basic theory and construction of informatization presents a rising trend. The management responsible for information technology in education is the Audio-Visual Education Center (or Centre for Educational Technology), computer center, network center, information center and other functions of the body. Due to the lack of uniform standards, the deviation in the understanding of the school and the university information technology education historical reasons, the team did not get enough attention in a number of institutions, even though the CIO system, IT governance and many new management concepts have been introduced in our country for a long time. These concepts do not get widely used in practice. In the aspect of evaluation of university information technology, especially information technology evaluation index system and evaluation theory and methods, research is still relatively not much. For the lack of scientific and systematic quantitative data, the study of such problems is still at an exploratory stage, and a recognized evaluation system is needed. Therefore, the number of studies on informatization management and evaluation which shows a horizontal state can be attributed to the separation of practice and theory.

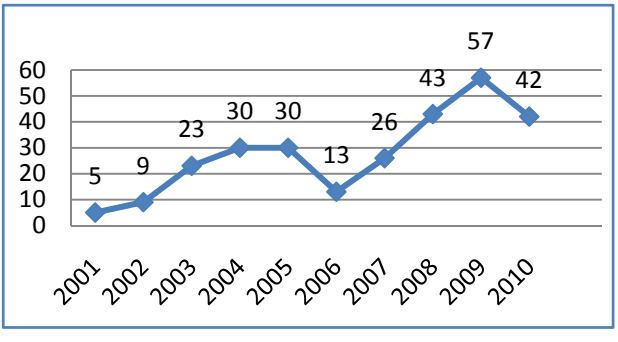


Fig. 1. The number of papers on higher education informatization from 2001 to 2010

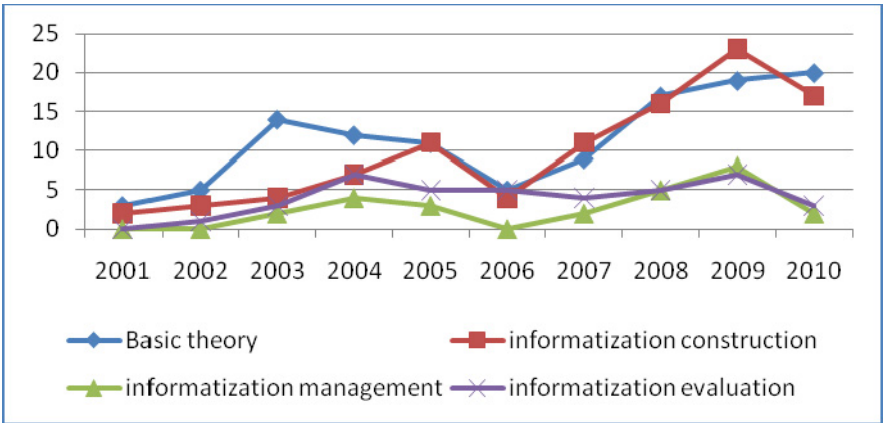


Fig. 2. The changing amount of papers on higher education informatization from 2001 to 2010

3.2 Research Subject Analysis

Through the statistical analysis of 20 articles (see Table 2) that selected from the highest frequency references from CNKI on higher education informatization, we can find that there are 5 articles written by Guodong Zhao, Dept of Educational technology in Graduate school of education of Peking university, and their citation frequency respectively is: 90, 59, 40, 40, 29. One of the 5 articles is about the current development research on informatization in the United States. Two articles are comparative studies on information development between China, the United States, and Japan (Hong Kong). Another two articles are on the current development research on informatization in China. Dongxing Jiang, from the center of Computer and information management in Tsinghua University, has published three articles about education informatization, and their frequency respectively is: 116, 38, and 33. The concept and construction plan of URP is put forward from his article "research on resource planning of university", and is the comprehensive, integrated, personalized, open, safe information system in practice primarily achieved in Tsinghua university, and provides other researched with a successful informatization integration pattern, so that it gets the highest quoted frequency (119) among all the related papers. There are three articles on higher education informatization written by the department of Information management in Nanjing University. Two articles of them are about Digital campus construction, and one is the study on Evaluation index system of higher education informatization. Their frequency respectively is: 51, 46 and 26.

We can learn that the research on informatization construction and practice in Tsinghua University, and the research on information development dynamic at home and abroad in Peking University are in the leading position. The information management department of Nanjing University also walks in the forefront of research field on higher education informatization.

Through the analysis of the articles' quantity (see Table 3) , it can be seen that some authors are much active, such as Dongxing Jiang from Tsinghua University, Yonggui Liu from Nanjing University, Guodong Zhao from Peking University, Chenghong Zhang from Fudan University.

3.3 Analysis of Research Content

The statistical result of Table 1 shows that the construction of application system, the current situation, trend and advice of the information development, informatization evaluation, informatization management, foreign experience, problems and countermeasures are hot topics. Here are some analyses about these six aspects.

(1) Application system construction

Informatization construction has different contents in different periods, and its contents are also changing [1]. There are three stages——single application, application integration and information integration——on the informatization construction in universities from the start to now [2].

Table 2. The 20 Most-Cited papers

No.	Paper Title	Author	cited times
1	University resource planning system	Dongxing Jiang, et al.	116
2	Comparative Research on ICT Implement in Higher Education of China& USA	Guodong Zhao	90
3	Investigation & analysis of the current status of information development in China higher education	Guodong Zhao; Yongzhong Huang	59
4	Building a New Generation Digital Campus of University	Xin Xu; Xinning Su	51
5	Conceptual Model of Universities in 21 Century	Lipeng Wan, et al.	46
6	Definition, framework and strategy of university digital campus based on knowledge management	Xibin Han, et al.	44
7	Comparative Study on the Campus Computing of China Mainland, the U.S.A and Hong Kong Special Administrative Region (HKSAR)	Guodong Zhao; Qiong Wang	40
8	Theoretic and Practical Thinking on E-learning	Guodong Zhao	40
9	Thinking about the informatization in high school	Min Shi	40
10	Planning and Practice of New Generation Digital Campus in Tsinghua University	Dongxing Jiang, et al.	40
11	A theoretic discussion on construction of informatization evaluation index system for China's high school	Baosheng An; Keding Zhong	38
12	Comprehensive evaluation on Higher education informatization level	Quanchao Zhao, et al.	37
13	Discussion about evaluation index system for higher education informatization	Junyue Liu, et al.	37
14	The Integration Method of Reflections and Practice in university informatization construction	Dongxing Jiang, et al.	33
15	The evaluation index system and methods of the education informatization in university and college.	Jun Fei, et al.	32
16	Investigation & analysis of the current status of informatization development in China higher education	Qiong Wang; Guodong Zhao	29
17	Reflection and inquiry on innovation of information pattern of higher education	Darou Chen	29
18	Discussion on Research Idea about the evaluation index system of the higher education informatization	Chenghong Zhang, et al	28
19	Research on Evaluating Indices System for IT in I-Campus	Jiahang Qin, et al.	26
20	Informatization of University Management and Management Team Construction	Huaijun Li	23

Table 3. The authors who published the most papers

Author's name	Dongxing Jiang	Yonggui Liu	Guodong Zhao	Chenghong Zhang
Author affiliationMain journals	China Education Info ; China Edu Info; Journal of Tsinghua University(Science and Technology)	Open Education Research; Modern Educational Technology	China Distance Education; China Educational Technology; Peking University Education Review	China Edu Info
Number of papers	6	5	4	3

For lack of overall planning, the stage of single application leads to the "information island" phenomenon. In order to eliminate this phenomenon, people began to emphasize the integration and conformity of application system. The management application system, such as electronic government affairs system for education, can solve the problems of informatization on management within the departments. The stage of application integration has the characteristics of information resources planning and comprehensive utilization of information, which emphasizes the exploiting, sharing and comprehensive utilization of information resources, in order to eliminate the phenomenon within the departments.

To solve the problems of lacking development planning in the construction process of higher education informatization, and the absence of resource sharing, and the difficulty of application integration, and the jumble of useless, Dongxing Jiang of Tsinghua university used the concept of enterprise ERP for reference, and put forward the URP (university resources plan) concept, and successfully applied it to the informatization practice of many universities [3].

(2) The current development situation, trend and advice of the informatization

The development of the informatization can be divided into theoretical research and empirical research. The empirical research is mainly the investigation including local survey and national survey. Guodong Zhao, et al, based on the National Education Informatization survey data of 2004, introduced the current development of informatization construction of domestic colleges and universities [4]. Xiaoyong Tian investigated by network, inspected and visited some investigated universities, and comprehensively analyzed some related data of higher education informatization development in Henan province in 2006 [5]. Jianguo Shi analyzed the construction and present application situation of higher education informatization in Zhejiang province by questionnaire and interview, and put forward the construction strategy to university education informatization in the whole province [6]; in 2008, the Ministry of Education organized a sampling survey for institutions of higher learning of 10 provinces、 cities and autonomous regions. In the aspect of theory, Kai Zheng

analyzed the informatization development track of each stage and present situation by the famous Nolan model, and predicted the development tendency according to the model [7]. Since the informatization construction in the college mainly relied on the network, and people could only get information passively from the network, Ning Zhang put forward the concept "mobile informatization" to help the mobile communication work in the university informatization construction [8]. Qintai Hu constructed the deep development framework of higher educational informationization construction from the angle of system engineering, and analyzed the future development trend of higher educational informatization [9].

(3) Informatization appraisal

Informatization evaluation is influenced by people's subjectivity and objective factors from levels of informatization, as well as the rapid development of the informatization, which may result in the lag of informatization evaluation system. So far, there isn't an authority standard of informatization evaluation system recognized by all the scholars.

Many academics have tried, and their researches focus on the model of informatization evaluation, indicator system and the weight calculation. The model based on GS Theory (gray theory), Fuzzy Theory, BSC (balance scorecard), and active degrees of Domain. The evaluation indicator is primarily constructed by Fuzzy Comprehensive Method. There are some methods such as Delphi Method and AHP for the calculation of weight.

With the analysis of evaluation system of university informatization in the domestic and foreign countries, and by referencing the issued index system of in China, Jiahang Qin used the indicator system of Chinese enterprise informatization for reference and built basic indicators and evaluation index of informatization evaluation indicator system for the university [10]. In addition, on account of resource perspective and process model, Chenghong Zhang explored the influence factors of university informatization from University Informatization Investment, Construction, Application and the whole Value generation process [11]. And by the use of SEM they got the three main factors affecting college informatization value production: finance, management ability of technology and information.

(4) Informatization management

Informatization management is a process during which information technology is fully used in the each link of management process. Information resources is widely developed, under the coordinating, planning and organizing of the organization in order to improve management efficiency, reduce management levels, promote an organizational flat structure, and reduce management cost[12].

In recent years, most university informatization management process has been achieved basically, which mainly shows in the aspects of infrastructure construction, information system construction, information resource integration and information personnel training. With the acceleration of informatization process in China, CIO management system be more and more valued, and has become a focus. Xiaomin Liu have made a detailed analysis and discussion on the origin and development of CIO, the development situation of CIO in the United States university, and the trial and

development of CIO construction system in China's universities[13]. To the situation that we have analogous organization to partly enforce the function of CIO, but there is not specific CIO post in university, Xiguo Fang proposes to straighten out complicated information management sections in the university, form a unified and authoritative comprehensive information administration section, and make this section play a part both in technology and in administration. He thinks it is a practical way of achieving CIO administration for university [14].

(5) Foreign experiences

The research on higher education informatization in China begins with learning the foreign advanced experiences. In the United States in 1990, Prof. Kenneth C. Green from Claremont University, first proposed "Campus Computing" (campus information) concept [15]. The number of empirical literature in 22 foreign countries reaches 13 to discuss the successful experience of the United States. Domestic scholars learn the successful experience of foreign countries through research, comparative studies, case studies, and international symposium held in a positive way. Guodong Zhao in 2002 based on the ACCS project survey data to compare China, the United States, and Hong Kong on the campus information technology achievements and problems in the recent years [16]. Jide Wang concluded from the United States, Britain, Japan and other developed countries in information technology in higher education, successful experiences were: attention of the government enterprises to participate in service outsourcing, the development of wireless networks, tilt to economically underdeveloped areas and vulnerable groups, emphasis information resource development, emphasis on teacher training and student information literacy training, and emphasis on education research and other information related to the theory [17]. Haixia Li and Ziyi Ju introduced the status and characteristics of the informatization in higher education of Korea [18]. Information on the study of foreign higher education ranges from information technology development status to the relevant organizations (such as the United States, EDUCAUSE, and professional associations), institutional mechanisms, teaching and research applications, in-depth into all areas of information technology.

(6) Problems and Countermeasures

After ten years of development, the higher education informatization in China has made great achievements. There are a range of issues, such as popularity, education equity issues, disclosure issues, network culture issues, ethical issues, information on pollution and safety issues, copyright issues, and lifelong education building problems. Haiqun Ma, et al introduced the legal basis for the University information disclosure policy, and discussed the information disclosure of China University of policy problems, and proposed solutions on issues related to recommendations [19]. Duyun Jiang pointed out that the information in the process of higher education had the issue of the network cultural impact and penetration of Western culture [20]. Li Duan, who pointed out that the information in the process of higher education had "weight does not respect others' bias", emphasized the process of information technology in colleges and universities should focus on human development [21].

4 Conclusion

The higher education informatization in China has experienced a decade of rapid growth and success, but there are some shortcomings. From the research method point of view, the higher education informatization research is mainly theoretical, and lack of the investigation studies, case studies and other research method. From the research perspective, the present study of higher education informatization starting from the material, the main research areas are infrastructure, information resources, qualified personnel, the funding situation, distance education, and information technology strategic planning, but few starting from the human. Teachers and students are lack of researches such as the utilization of digital campus which costs lots of money, the level of use, the match of demand and the actual, the satisfaction of the school of information technology by the teachers and students. Personnel training, after all, are the focus of colleges and universities. The university is in the social environment, not only relying on the existence of society, but also serving the community. Personnel training are ultimately for social development. University information system development and socio-economic development should be closely linked. What kind of information professionals is the needs of the community. This should also be a major issue of the higher education informatization technology researches in future.

References

1. Song, J., Chen, S.: The Enlightenment of Informatization Development Stages Theory to the Informatization in Our Colleges and Universities. China Education Info. (5), 8–10 (2007)
2. Yang, L.: Research and application of Data Integration Platform in Universities Informatization. Southeast University, Nanjing (2009)
3. Jiang, D., et al.: University resource planning system. Journal of Tsinghua University (Sci. &Tech.) 44(4), 572–576 (2004)
4. Zhao, G., Huang, Y.: Investigation & analysis of the current status of information development in China higher education. China Distance Education (8), 43–48 (2005)
5. Tian, X., Liang, C., Fan, S., et al.: The annual report 2006: Higher Education Informatization of Henan Province. China Educational Technology (12), 41–44 (2007)
6. Shi, J., Xiang, X., Cheng, L.: Status Quo and Countermeasure of Higher Education Informatization in Zhejiang Province. China Educational Technology (4), 32–35 (2010)
7. Zheng, K., Nie, R.: Analysis on the Status Quo and Trends of Universities' Informatization Development Based on "Nolan" model. China Education Info. (21), 13–15 (2009)
8. Zhang, N., Zhu, J.: The Discussion of Mobile Information Construction in University. Computer Knowledge and Technology 5(25), 7263–7265 (2009)
9. Hu, Q.: The Deep Development Framework and Trend Analysis of Higher Educational Informationization. Educational Research (10), 97–101 (2009)
10. Qin, J., Xu, X., Su, X.: Research on Evaluating Indices System for IT in I-Campus. New Technology of Library and Information Service (4), 63–69 (2006)
11. Zhang, C., Hu, Y., Zhang, C.: Influence Factors and empirical research on university informatization based on Resource-based view. Journal of Fudan University (Science Edition) 48(2), 271–276 (2009)

12. Lang, D.: The Analysis on Three-characteristic of Boosting Informatization Management of Undergraduates. Central China Normal University, Wuhan (2006)
13. Liu, X., Chen, A.: A Study on Establishing University CIO System during Higher Education Informatization. Open Education Research 11(2), 42–47 (2005)
14. Fang, X.: The Reality Ways of university CIO management in China. China Education Info. (1), 16–17 (2011)
15. Zhang, W.: Overview on Higher Education Informatization Research in the last decade of China. Liaoning Education Research (8), 49–51 (2007)
16. Zhao, G., Wang, Q.: Comparative Study on the Campus Computing of China Mainland, the U.S.A and Hong Kong Special Administrative Region (HKSAR). Modern Distance Education (2), 12–18 (2003)
17. Wang, J., Han, M.: The experience and enlightenment of higher education Informatization from Developed countries. China Educational Technology (6), 90–94 (2005)
18. Li, H., Ju, Z.: The Development and Revelation of the Informationalization in Higher Education of Korea. Comparative Education Review (7), 30–34 (2009)
19. Ma, H., Wen, L.: The Analysis of the Legal Origin and Problems about University Information Disclosure Policy of China. Journal of Modern Information 31(3), 3–6 (2011)
20. Jiang, D.: The Cultural Question in the Informatization of Higher Education and How to Deal with it. Journal of Zhengzhou University (Philosophy and Social Science Edition) 36(2), 79–82 (2003)
21. Duan, L., Yao, L.: Focus on human development in the process of higher education Informatization. China Distance Education (7), 18–20 (2003)

Author Index