LNCS 6774

Randall Shumaker (Ed.)

Virtual and Mixed Reality – Systems and Applications

International Conference, Virtual and Mixed Reality 2011
Held as Part of HCI International 2011
Orlando, FL, USA, July 2011, Proceedings, Part II

2 Part II

HCI2011
INTERNATIONAL

Springer

Lecture Notes in Computer Science 6774

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

Randall Shumaker (Ed.)

Virtual and Mixed Reality – Systems and Applications

International Conference, Virtual and Mixed Reality 2011
Held as Part of HCI International 2011
Orlando, FL, USA, July 9-14, 2011
Proceedings, Part II

Springer

Volume Editor

Randall Shumaker
University of Central Florida
Institute for Simulation and Training
3100 Technology Parkway and 3280 Progress Drive
Orlando, FL 32826, USA
E-mail: shumaker@ist.ucf.edu

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-22023-4 e-ISBN 978-3-642-22024-1
DOI 10.1007/978-3-642-22024-1
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: Applied for

CR Subject Classification (1998): H.5, H.4, I.3, I.2, C.3, I.4, I.6

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

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)

Foreword

The 14th International Conference on Human–Computer Interaction, HCI International 2011, was held in Orlando, Florida, USA, July 9–14, 2011, jointly with the Symposium on Human Interface (Japan) 2011, the 9th International Conference on Engineering Psychology and Cognitive Ergonomics, the 6th International Conference on Universal Access in Human–Computer Interaction, the 4th International Conference on Virtual and Mixed Reality, the 4th International Conference on Internationalization, Design and Global Development, the 4th International Conference on Online Communities and Social Computing, the 6th International Conference on Augmented Cognition, the Third International Conference on Digital Human Modeling, the Second International Conference on Human-Centered Design, and the First International Conference on Design, User Experience, and Usability.

A total of 4,039 individuals from academia, research institutes, industry and governmental agencies from 67 countries submitted contributions, and 1,318 papers that were judged to be of high scientific quality were included in the program. These papers address the latest research and development efforts and highlight the human aspects of design and use of computing systems. The papers accepted for presentation thoroughly cover the entire field of human–computer interaction, addressing major advances in knowledge and effective use of computers in a variety of application areas.

This volume, edited by Randall Shumaker, contains papers in the thematic area of virtual and mixed reality (VMR), addressing the following major topics:

- VR in education, training and health
- VR for culture and entertainment
- Virtual humans and avatars
- Developing virtual and mixed environments

The remaining volumes of the HCI International 2011 Proceedings are:

- Volume 1, LNCS 6761, Human–Computer Interaction—Design and Development Approaches (Part I), edited by Julie A. Jacko
- Volume 2, LNCS 6762, Human–Computer Interaction—Interaction Techniques and Environments (Part II), edited by Julie A. Jacko
- Volume 3, LNCS 6763, Human–Computer Interaction—Towards Mobile and Intelligent Interaction Environments (Part III), edited by Julie A. Jacko
- Volume 4, LNCS 6764, Human–Computer Interaction—Users and Applications (Part IV), edited by Julie A. Jacko
- Volume 5, LNCS 6765, Universal Access in Human–Computer Interaction—Design for All and eInclusion (Part I), edited by Constantine Stephanidis
- Volume 6, LNCS 6766, Universal Access in Human–Computer Interaction—Users Diversity (Part II), edited by Constantine Stephanidis

- Volume 7, LNCS 6767, Universal Access in Human–Computer Interaction—Context Diversity (Part III), edited by Constantine Stephanidis
- Volume 8, LNCS 6768, Universal Access in Human–Computer Interaction—Applications and Services (Part IV), edited by Constantine Stephanidis
- Volume 9, LNCS 6769, Design, User Experience, and Usability—Theory, Methods, Tools and Practice (Part I), edited by Aaron Marcus
- Volume 10, LNCS 6770, Design, User Experience, and Usability—Understanding the User Experience (Part II), edited by Aaron Marcus
- Volume 11, LNCS 6771, Human Interface and the Management of Information—Design and Interaction (Part I), edited by Michael J. Smith and Gavriel Salvendy
- Volume 12, LNCS 6772, Human Interface and the Management of Information—Interacting with Information (Part II), edited by Gavriel Salvendy and Michael J. Smith
- Volume 13, LNCS 6773, Virtual and Mixed Reality—New Trends (Part I), edited by Randall Shumaker
- Volume 15, LNCS 6775, Internationalization, Design and Global Development, edited by P.L. Patrick Rau
- Volume 16, LNCS 6776, Human-Centered Design, edited by Masaaki Kurosu
- Volume 17, LNCS 6777, Digital Human Modeling, edited by Vincent G. Duffy
- Volume 18, LNCS 6778, Online Communities and Social Computing, edited by A. Ant Ozok and Panayiotis Zaphiris
- Volume 19, LNCS 6779, Ergonomics and Health Aspects of Work with Computers, edited by Michelle M. Robertson
- Volume 20, LNAI 6780, Foundations of Augmented Cognition: Directing the Future of Adaptive Systems, edited by Dylan D. Schmorrow and Cali M. Fidopiastis
- Volume 21, LNAI 6781, Engineering Psychology and Cognitive Ergonomics, edited by Don Harris
- Volume 22, CCIS 173, HCI International 2011 Posters Proceedings (Part I), edited by Constantine Stephanidis
- Volume 23, CCIS 174, HCI International 2011 Posters Proceedings (Part II), edited by Constantine Stephanidis

I would like to thank the Program Chairs and the members of the Program Boards of all Thematic Areas, listed herein, for their contribution to the highest scientific quality and the overall success of the HCI International 2011 Conference.

In addition to the members of the Program Boards, I also wish to thank the following volunteer external reviewers: Roman Vilimek from Germany, Ramalingam Ponnusamy from India, Si Jung “Jun” Kim from the USA, and Ilia Adami, Iosif Klironomos, Vassilis Kouroumalis, George Margetis, and Stavroula Ntoa from Greece.

This conference would not have been possible without the continuous support and advice of the Conference Scientific Advisor, Gavriel Salvendy, as well as the dedicated work and outstanding efforts of the Communications and Exhibition Chair and Editor of HCI International News, Abbas Moallem.

I would also like to thank for their contribution toward the organization of the HCI International 2011 Conference the members of the Human–Computer Interaction Laboratory of ICS-FORTH, and in particular Margherita Antona, George Paparoulis, Maria Pitsoulaki, Stavroula Ntoa, Maria Bouhli and George Kapnas.

July 2011 Constantine Stephanidis

Organization

Ergonomics and Health Aspects of Work with Computers

Program Chair: Michelle M. Robertson

Arne Aarås, Norway
Pascale Carayon, USA
Jason Devereux, UK
Wolfgang Friesdorf, Germany
Martin Helander, Singapore
Ed Israelski, USA
Ben-Tzion Karsh, USA
Waldemar Karwowski, USA
Peter Kern, Germany
Danuta Koradecka, Poland
Nancy Larson, USA
Kari Lindström, Finland
Brenda Lobb, New Zealand
Holger Luczak, Germany
William S. Marras, USA
Aura C. Matias, Philippines
Matthias Rötting, Germany
Michelle L. Rogers, USA
Dominique L. Scapin, France
Lawrence M. Schleifer, USA
Michael J. Smith, USA
Naomi Swanson, USA
Peter Vink, The Netherlands
John Wilson, UK

Human Interface and the Management of Information

Program Chair: Michael J. Smith

Hans-Jörg Bullinger, Germany
Alan Chan, Hong Kong
Shin'ichi Fukuzumi, Japan
Jon R. Gunderson, USA
Michitaka Hirose, Japan
Jhilmil Jain, USA
Yasufumi Kume, Japan
Mark Lehto, USA
Hirohiko Mori, Japan
Fiona Fui-Hoon Nah, USA
Shogo Nishida, Japan
Robert Proctor, USA
Youngho Rhee, Korea
Anxo Cereijo Roibás, UK
Katsunori Shimohara, Japan
Dieter Spath, Germany
Tsutomu Tabe, Japan
Alvaro D. Taveira, USA
Kim-Phuong L. Vu, USA
Tomio Watanabe, Japan
Sakae Yamamoto, Japan
Hidekazu Yoshikawa, Japan
Li Zheng, P. R. China

Human–Computer Interaction

Program Chair: Julie A. Jacko

Sebastiano Bagnara, Italy
Sherry Y. Chen, UK
Marvin J. Dainoff, USA
Jianming Dong, USA
John Eklund, Australia
Xiaowen Fang, USA
Ayse Gurses, USA
Vicki L. Hanson, UK
Sheue-Ling Hwang, Taiwan
Wonil Hwang, Korea
Yong Gu Ji, Korea
Steven A. Landry, USA
Gitte Lindgaard, Canada
Chen Ling, USA
Yan Liu, USA
Chang S. Nam, USA
Celestine A. Ntuen, USA
Philippe Palanque, France
P.L. Patrick Rau, P.R. China
Ling Rothrock, USA
Guangfeng Song, USA
Steffen Staab, Germany
Wan Chul Yoon, Korea
Wenli Zhu, P.R. China

Engineering Psychology and Cognitive Ergonomics

Program Chair: Don Harris

Guy A. Boy, USA
Pietro Carlo Cacciabue, Italy
John Huddlestone, UK
Kenji Itoh, Japan
Hung-Sying Jing, Taiwan
Wen-Chin Li, Taiwan
James T. Luxhøj, USA
Nicolas Marmaras, Greece
Sundaram Narayanan, USA
Mark A. Neerincx, The Netherlands
Jan M. Noyes, UK
Kjell Ohlsson, Sweden
Axel Schulte, Germany
Sarah C. Sharples, UK
Neville A. Stanton, UK
Xianghong Sun, P.R. China
Andrew Thatcher, South Africa
Matthew J.W. Thomas, Australia
Mark Young, UK
Rolf Zon, The Netherlands

Universal Access in Human–Computer Interaction

Program Chair: Constantine Stephanidis

Julio Abascal, Spain
Ray Adams, UK
Elisabeth André, Germany
Margherita Antona, Greece
Chieko Asakawa, Japan
Christian Bühler, Germany
Jerzy Charytonowicz, Poland
Pier Luigi Emiliani, Italy
Michael Fairhurst, UK
Dimitris Grammenos, Greece
Andreas Holzinger, Austria
Simeon Keates, Denmark
Georgios Kouroupetroglou, Greece
Sri Kurniawan, USA
Patrick M. Langdon, UK
Seongil Lee, Korea

Zhengjie Liu, P.R. China
Klaus Miesenberger, Austria
Helen Petrie, UK
Michael Pieper, Germany
Anthony Savidis, Greece
Andrew Sears, USA
Christian Stary, Austria
Hirotada Ueda, Japan
Jean Vanderdonckt, Belgium
Gregg C. Vanderheiden, USA
Gerhard Weber, Germany
Harald Weber, Germany
Panayiotis Zaphiris, Cyprus

Virtual and Mixed Reality

Program Chair: Randall Shumaker

Pat Banerjee, USA
Mark Billinghurst, New Zealand
Charles E. Hughes, USA
Simon Julier, UK
David Kaber, USA
Hirokazu Kato, Japan
Robert S. Kennedy, USA
Young J. Kim, Korea
Ben Lawson, USA
Gordon McK Mair, UK
David Pratt, UK
Albert “Skip” Rizzo, USA
Lawrence Rosenblum, USA
Jose San Martin, Spain
Dieter Schmalstieg, Austria
Dylan Schmorrow, USA
Kay Stanney, USA
Janet Weisenford, USA
Mark Wiederhold, USA

Internationalization, Design and Global Development

Program Chair: P.L. Patrick Rau

Michael L. Best, USA
Alan Chan, Hong Kong
Lin-Lin Chen, Taiwan
Andy M. Dearden, UK
Susan M. Dray, USA
Henry Been-Lirn Duh, Singapore
Vanessa Evers, The Netherlands
Paul Fu, USA
Emilie Gould, USA
Sung H. Han, Korea
Veikko Ikonen, Finland
Toshikazu Kato, Japan
Esin Kiris, USA
Apala Lahiri Chavan, India
James R. Lewis, USA
James J.W. Lin, USA
Rungtai Lin, Taiwan
Zhengjie Liu, P.R. China
Aaron Marcus, USA
Allen E. Milewski, USA
Katsuhiko Ogawa, Japan
Oguzhan Ozcan, Turkey
Girish Prabhu, India
Kerstin Röse, Germany
Supriya Singh, Australia
Alvin W. Yeo, Malaysia
Hsiu-Ping Yueh, Taiwan

Online Communities and Social Computing

Program Chairs: A. Ant Ozok, Panayiotis Zaphiris

Chadia N. Abras, USA
Chee Siang Ang, UK
Peter Day, UK
Fiorella De Cindio, Italy
Heidi Feng, USA
Anita Komlodi, USA
Piet A.M. Kommers, The Netherlands
Andrew Laghos, Cyprus
Stefanie Lindstaedt, Austria
Gabriele Meiselwitz, USA
Hideyuki Nakanishi, Japan
Anthony F. Norcio, USA
Ulrike Pfeil, UK
Elaine M. Raybourn, USA
Douglas Schuler, USA
Gilson Schwartz, Brazil
Laura Slaughter, Norway
Sergei Stafeev, Russia
Asimina Vasalou, UK
June Wei, USA
Haibin Zhu, Canada

Augmented Cognition

Program Chairs: Dylan D. Schmorrow, Cali M. Fidopiastis

Monique Beaudoin, USA
Chris Berka, USA
Joseph Cohn, USA
Martha E. Crosby, USA
Julie Drexler, USA
Ivy Estabrooke, USA
Chris Forsythe, USA
Wai Tat Fu, USA
Marc Grootjen, The Netherlands
Jefferson Grubb, USA
Santosh Mathan, USA
Rob Matthews, Australia
Dennis McBride, USA
Eric Muth, USA
Mark A. Neerincx, The Netherlands
Denise Nicholson, USA
Banu Onaral, USA
Kay Stanney, USA
Roy Stripling, USA
Rob Taylor, UK
Karl van Orden, USA

Digital Human Modeling

Program Chair: Vincent G. Duffy

Karim Abdel-Malek, USA
Giuseppe Andreoni, Italy
Thomas J. Armstrong, USA
Norman I. Badler, USA
Fethi Calisir, Turkey
Daniel Carruth, USA
Keith Case, UK
Julie Charland, Canada
Yaobin Chen, USA
Kathryn Cormican, Ireland
Daniel A. DeLaurentis, USA
Yingzi Du, USA
Okan Ersoy, USA
Enda Fallon, Ireland
Yan Fu, P.R. China
Afzal Godil, USA

Ravindra Goonetilleke, Hong Kong
Anand Gramopadhye, USA
Lars Hanson, Sweden
Pheng Ann Heng, Hong Kong
Bo Hoege, Germany
Hongwei Hsiao, USA
Tianzi Jiang, P.R. China
Nan Kong, USA
Steven A. Landry, USA
Kang Li, USA
Zhizhong Li, P.R. China
Tim Marler, USA
Ahmet F. Ozok, Turkey
Srinivas Peeta, USA
Sudhakar Rajulu, USA
Matthias Rötting, Germany
Matthew Reed, USA
Johan Stahre, Sweden
Mao-Jiun Wang, Taiwan
Xuguang Wang, France
Jingzhou (James) Yang, USA
Gulcin Yucel, Turkey
Tingshao Zhu, P.R. China

Human-Centered Design

Program Chair: Masaaki Kurosu

Julio Abascal, Spain
Simone Barbosa, Brazil
Tomas Berns, Sweden
Nigel Bevan, UK
Torkil Clemmensen, Denmark
Susan M. Dray, USA
Vanessa Evers, The Netherlands
Xiaolan Fu, P.R. China
Yasuhiro Horibe, Japan
Jason Huang, P.R. China
Minna Isomursu, Finland
Timo Jokela, Finland
Mitsuhiko Karashima, Japan
Tadashi Kobayashi, Japan
Seongil Lee, Korea
Kee Yong Lim, Singapore
Zhengjie Liu, P.R. China
Loïc Martínez-Normand, Spain
Monique Noirhomme-Fraiture, Belgium
Philippe Palanque, France
Annelise Mark Pejtersen, Denmark
Kerstin Röse, Germany
Dominique L. Scapin, France
Haruhiko Urokohara, Japan
Gerrit C. van der Veer, The Netherlands
Janet Wesson, South Africa
Toshiki Yamaoka, Japan
Kazuhiko Yamazaki, Japan
Silvia Zimmermann, Switzerland

Design, User Experience, and Usability

Program Chair: Aaron Marcus

Ronald Baecker, Canada
Barbara Ballard, USA
Konrad Baumann, Austria
Arne Berger, Germany
Randolph Bias, USA
Jamie Blustein, Canada
Ana Boa-Ventura, USA
Lorenzo Cantoni, Switzerland
Sameer Chavan, Korea
Wei Ding, USA
Maximilian Eibl, Germany
Zelda Harrison, USA

Rüdiger Heimgärtner, Germany
Brigitte Herrmann, Germany
Sabine Kabel-Eckes, USA
Kaleem Khan, Canada
Jonathan Kies, USA
Jon Kolko, USA
Helga Letowt-Vorbek, South Africa
James Lin, USA
Frazer McKimm, Ireland
Michael Renner, Switzerland
Christine Ronnewinkel, Germany
Elizabeth Rosenzweig, USA
Paul Sherman, USA
Ben Shneiderman, USA
Christian Sturm, Germany
Brian Sullivan, USA
Jaakko Villa, Finland
Michele Visciola, Italy
Susan Weinschenk, USA

HCI International 2013

The 15th International Conference on Human–Computer Interaction, HCI International 2013, will be held jointly with the affiliated conferences in the summer of 2013. It will cover a broad spectrum of themes related to human–computer interaction (HCI), including theoretical issues, methods, tools, processes and case studies in HCI design, as well as novel interaction techniques, interfaces and applications. The proceedings will be published by Springer. More information about the topics, as well as the venue and dates of the conference, will be announced through the HCI International Conference series website: http://www.hci-international.org/

General Chair
Professor Constantine Stephanidis
University of Crete and ICS-FORTH
Heraklion, Crete, Greece
Email: cs@ics.forth.gr

Table of Contents – Part II

Part I: VR in Education, Training and Health

Part II: VR for Culture and Entertainment

Part III: Virtual Humans and Avatars

Part IV: Developing Virtual and Mixed Environments

Table of Contents – Part I

Part I: Augmented Reality Applications

Part II: Virtual and Immersive Environments

Part III: Novel Interaction Devices and Techniques in VR

Part IV: Human Physiology and Behaviour in VR Environments

Part I
VR in Education, Training and Health

Serious Games for Psychological Health Education

Anya Andrews

Instructional Systems Architect, Senior Research Scientist
Institute for Simulation and Training, University of Central Florida
3100 Technology Parkway, Orlando FL, 32826
aandrews@ist.ucf.edu

Abstract. This paper presents a summary of recent research efforts aiming to address modern psychological health education needs through the use of innovative instructional tools. The current body of research on virtual learning environments and serious games as they relate to psychological treatment shows promising results, especially in the case of the instructional interventions that provide an optimal blend of education and training and focus on the psychological health knowledge acquisition as well as appropriate stress management skills and behaviors. In concert with the theoretical and research foundations within the psychological health domain and pedagogical precepts in the area of simulation and game-based learning, this article also presents design considerations for serious games for psychological health.

Keywords: Serious Games, Psychological Health Education, Mental Health, Virtual Learning Environments.

1 Research Overview

The fundamental goal of the Research and Development (R&D) effort highlighted in this paper was to conceptualize innovative solutions to educate service members about the military Psychological Health (PH) process and equip them with the skills necessary to ensure their psychological well-being at every stage of the military deployment cycle. The research and analysis activities conducted during this effort were aimed to inform the design of an interactive instructional system for psychological health education. The key areas under investigation included the psychological issues commonly encountered within military environments, the coping strategies that can be employed by service members to mitigate those issues, available care options within the military PH process, and associated barriers to care.

During the course of the research effort, an extensive review of academic and military publications was performed to identify and analyze the theoretical foundations and empirical evidence for explaining the relationship between PH factors, available care options, coping strategies, and stress management skills relevant to military environments. Qualitative meta-analysis techniques were used to analyze and confirm/resolve apparent consensus and contradictions within the research findings by exploring the relations between study characteristics and findings. A rich knowledge base of first-hand perspectives from military personnel and their families was assembled from interactive interviews with military experts, their family members as well as support group websites,

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 3–10, 2011.

social networking sites, and blogosphere. This information correlated with the existing body of research and provided a collection of storyline vignettes to be leveraged for the design and development of the PH instructional interventions to be delivered to the learners through the use of serious games.

2 Pedagogical Foundations for Military Psychological Health Education

The research links the absence of effective mental health training and education programs to the lack of critical coping skills that could safeguard “at risk” populations against potential psychological health issues. Focused on providing relevant skills, such as recognizing the signs and symptoms of stress, identifying sources of stress, applying relevant coping strategies, and seeking professional help and/or social support, these programs can help prevent psychological disturbances resulting from traumatic stress events, minimize the severity of existing mental health disorders, and ultimately protect the psychological well-being of our troops and their families. It is interesting to note that most of the existing research focuses on mitigating clinical conditions, such as PTSD for the post-deployment service members.

Resilience is frequently identified as an important competency for prevention of PTSD. The Battlemind program does offer pre-deployment resilience training, although it does not specifically address mental health issues and prevention strategies. While there are a few resilience programs that address some of the mental health issues for combat readiness; there is currently no system-wide military resilience training strategy to mentally prepare service members for the stresses of combat [1]. During the meta-analysis, the research team determined that most of the available resilience training resources were not necessarily affiliated with military organizations. It is also unknown what percentage of service members are aware of such programs.

Although the current research appears to be limited in terms of assessing the power of prevention, there is an obvious consensus about the importance of awareness-oriented interventions that can be introduced at the pre-deployment phase, thereby, giving service members the tools to help mitigate potential psychological challenges. Research focused on the specific mental health needs of various “at risk” populations, barriers to accessing care, and the efficacy of existing prevention and intervention programs is critical to making mental health care more relevant, available, and effective. In the training programs, specific emphasis must be placed on the development of the high-impact skills, such as resilience, self-regulation, problem identification, self-reflection, strategy formation, self-adaptation, and others.

3 Prevention as a Pedagogical Construct

The power of prevention versus treatment has been extensively recognized in general psychology literature, however, military psychology tends to be largely focused on the after-effects of traumatic combat experiences, with relatively low emphasis on prevention. This tendency can easily be associated with the lack of awareness among service members about the military stress prevention programs in general. While military psychology research does acknowledge prevention in the form of *pre*-deployment preparation and training, there are mixed opinions about whether more serious conditions, such as PTSD, can be prevented.

Although the value of prevention has been recognized by research and there is an obvious trend towards addressing the psychological needs of military personnel starting at the pre-deployment stage, the system-wide focus is still largely placed on dealing with post-deployment after effects. At the same time, in response to the frightening suicide rates among military personnel resulting from long deployments, continuous separations from family, and the perceived stigma associated with seeking PH help, numerous military organizations initiated a series of training and education efforts for military suicide prevention and stigma-related issues.

What is prevention? Feldner, Monson, & Friedman [2] emphasized that prevention exclusively focuses on reducing the *incidence* (rather than prevalence) of a disorder. Thus, three categories of prevention approaches have been distinguished:

- A *universal intervention* is applied to all members of the population, regardless of their risk for developing a disorder.
- A *selective intervention* targets only persons at risk for developing, but showing no signs of, a disorder.
- An *indicated intervention* is aimed at individuals demonstrating aspects of a disorder but who are sub-syndromal.

According to Cannon-Bowers & Bowers [3], prevention training for populations with a potentially high risk of PTSD will significantly help minimize risk, minimize cost, minimize impact on healthcare system, and overall improve the quality of life. Widely described in literature and empirically evaluated, Cognitive Behavioral Therapy (CBT) is an effective method used for treatment and prevention of mental health disorders. Feldner et al [2] report promising results from a series of studies using CBT in comparison to repeated assessment, delayed assessment, and supportive counseling, collectively resulting in significantly PTSD rates. In the quest to find novel approaches to facilitate prevention of mental health disorders, Stetz et al [4] conducted a series of experiments using Virtual Reality Stress Inoculation Training (VR-SIT) and Coping Training (CT) via relaxation behavior techniques and found evidence of the 'hardening' effect against combat stress." [4]. While this study blazes a trail for the use of virtual environments for stress prevention interventions, the absence of long-term data for the participants' ability to cope with their traumatic exposure requires further investigation and validation of the effectiveness of these two treatment techniques for prevention purposes.

4 Virtual Learning Environments and Serious Games for as Psychological Health Education Tools

The current body of research surrounding virtual learning environments and serious games as they relate to psychological treatment shows promising results, especially in the case of the training interventions that provide an optimal blend of education and training and focus on the PH knowledge acquisition as well as stress management skills and behaviors. The notion of Serious Games (SG) refers to the use of state-of-the-art gaming technologies for developing adaptive instructional solutions. According to the Serious Games Forum (2008), "*a serious game may be a simulation which has the look and feel of a game, but corresponds to non-game events or processes, including business operations and military operations*." An Immersive Learning Simulation (ILS), also known as a Serious Game, is "*an optimized blend of simulation, game element, and pedagogy that leads to the learner being motivated by, and immersed into, the purpose and goals of a learning interaction*" [5].

Serious games use meaningful contextualization, and optimized experience, to successfully integrate the engagement of well-designed games with serious learning goals [6]. Gaming environments offer exceptional potential for teaching cognitive and behavioral skills by providing opportunities for simulating the real-life situations and conditions, under which the development of these skills occurs. Games motivate learners to take responsibility for their own learning, which leads to intrinsic motivation contained by the game-based learning method itself. While engaged in a serious game, the learners tend to retain a significant amount of information by building cognitive maps. This process fosters adaptive decision making, a critical competency of modern military personnel.

4.1 Virtual Reality Exposure Training (VRET)

Virtual Reality Exposure Training (VRET) is a growing treatment protocol for specific anxieties, phobias, panic disorders, and PTSD [7]. The theory behind exposure-based training maintains that repeated exposure to fearful situations within a safe and controlled environment provides unique opportunities for the user to rehearse a different emotional reaction. Successful rehearsals activate new brain patterns in response to the emotional situation. Ultimately, repeated successful exposure may enhance emotional regulation in live situations.

In practice, exposure training is typically a therapeutic intervention administered in the presence of a clinician verbally guiding the person with a psychological disorder to imagine, narrate, and emotionally process the traumatic event. Although this type of imagined exposure training has been proven effective for PTSD, there is an inherent challenge with this visualization method, because one of the cardinal indicators of PTSD is avoidance of the feared stimuli. Many patients are unable or unwilling to visualize the feared content in their imaginations [8]. VRET provides a way to circumvent this avoidance tendency by immersing the user into a computer-generated environment that systematically exposes him to the fear stimuli. Using this methodology, the University of Southern California, Institute for Creative Technologies (ICT) developed a VRET system for combat-related PTSD. *Virtual Iraq* and *Virtual Afghanistan* provide relevant contexts for exposure therapy, such as a middle-eastern themed city and desert road environments. In addition to visual stimuli, the program can deliver directional 3D audio, olfactory, and tactile stimuli. A separate interface allows a clinician to control the situation in real time and customize the scenario to meet the needs of a specific patient. The initial clinical trial outcomes of this approach appear to be encouraging.

4.2 Second Life as a Healing Space for Veterans

Another notable example of virtual environment to help soldiers reintegrate into civilian life revolves around an exclusive space for veterans within the online virtual world, Second Life®. This social networking outlet provides a safe place for soldiers to find camaraderie, become aware of different treatments options, and seek the help that they need. This virtual world environment serves as a healing space for veterans upon re-entry [9]. As mentioned earlier, numerous barriers may prevent veterans from seeking mental healthcare upon their return from war. Many soldiers are geographically dispersed and find it difficult to create a local community of returning veterans. The leading-edge virtual immersive therapies are not widespread and are not available to most soldiers. Using the online social site such as Second Life® circumvents these issues by providing a place where veterans can interact anonymously within the site from the privacy of their own homes. This social space provides a network for veterans to find companionship with other soldiers who have recently returned from war. The site also provides opportunities for veterans to practice a number of Complementary and Alternative Medicine (CAM)-based exercises, such as relaxation techniques, cognitive behavioral therapy, and self-reflection. A resources area provides access to websites and information on which traditional therapies a person might find beneficial. Virtual human agents are available within the site to provide guidance and resource information.

4.3 "Walk in My Shoes" – A Serious Game for Psychological Health

Within the context of the described effort, the research team designed a serious game called "Walk in My Shoes", which is intended for a diverse learner audience, ranging from military personnel at different stages of deployment cycle (pre-deployment, deployment/in-theatre, and post-deployment) to veteran and civilian populations whose psychological well-being may be affected by the stress-inducing challenges of the military life. As a result of a series of analyses, the research team determined that the serious game approach driven by the guided exploratory learning strategy would represent an optimal way to foster PH knowledge and skill acquisition in an engaging self-reinforcing environment. The game is designed to introduce the learner to the common psychological health challenges faced by military personnel today. The user will explore different coping strategies and help options available at different stages of the military deployment cycle. Focused on the military PH process intends not only to raise the learner awareness regarding the PH-related issues and available options, but also help develop stress prevention, recognition, and management skills.

The instructional strategy of the serious game "Walk in My Shoes" is based on the narrative-centric guided exploratory approach and the best practices of digital storytelling. Due to the motivational power of narrative, digital storytelling can be both engaging and effective especially in discovery learning that lends itself to the learners' active exploration of the subject matter. In contrast to didactic pedagogies based on memorization of facts, discovery learning encourages students to learn by posing questions and answering them through analysis, reflection, or problem-solving activities. Narrative-centered learning environments offer significant potential for supporting guided exploratory learning. By taking advantage of the inherent structure of narrative, narrative-centered environments provide learners with engaging worlds in which they actively participate in motivating story-based problem-solving activities. Using digital storytelling techniques, meaningful scenarios encapsulated within a storyline-driven mini-virtual world will illustrate a full range of psychological issue cases commonly encountered in military environments.

As a narrative-centered serious game, “Walk in My Shoes” takes the learners on a guided exploratory journey within a purposefully constructed virtual mini-world where the learners can face the psychological health challenges common within military environment and practice a variety of coping strategies as they “walk in the shoes” of the key game characters. The learner enters the game through a virtual living room and “meets” the key game characters via a Photo Album located on the coffee table. The photographs in the album depict individual characters facing different sets of challenges and issues typical for a particular stage of the deployment cycle. The learner will have an opportunity to “walk in the shoes” of each character and experience a variety of psychological challenges as well as explore appropriate or less appropriate coping strategies and help options. A virtual mentor, Dr. Livewell is a trained military psychologist who provides valuable tips and advice along the way.

The prototype development effort was primarily focused on the pre-deployment stage and included the learning activities designed to address the challenges of pre-deployment preparations. Thus, the virtual mini-world included a number of environments (e.g. legal aid office, bank, parents’ house, friend’s house, Family Readiness Group meeting building, park area, bar, and others), offering the opportunities to engage in meaningful interactions with other characters who would provide the learner with a wealth of pre-deployment-related information presented using engaging digital storytelling techniques. During the “journey”, the learner receives information from a variety of channels including legal/financial professionals, family, friends, support group members, etc. The auxiliary game characters will play a crucial role in raising the learner’s awareness about things that need to be done to gain the “peace of mind” at the pre-deployment stage.

In addition to raising the learner’s awareness, the game also provides opportunities for the learner to engage in decision-making activities and cognitive exercises. Along with collecting useful tips and advice from the strategic characters, the game offers the means to reflect upon one’s thoughts and feelings and practice applying a number of coping strategies (e.g. identifying one’s own psychological vulnerabilities, cognitive restructuring/reframing, thought distortion removal, setting expectations, etc.) to deal with the psychological burden of the pre-deployment stage. A variety of cognitive exercises for practicing the coping strategies are woven into the storyline in the form of a fictitious Mood Log and a number of mini-games associated with the events unfolding within individual environments of the virtual mini-world.

4.4 Games for Psychological Health: Key Design Recommendations

The results of the described effort are both theoretical and practical in nature. The research and development outcomes contribute to the following major bodies of research: psychological health, military health education, and game-based learning. In concert with theoretical underpinnings for game-based instruction and the author’s experience with serious games for mental health, the following design recommendations are offered to educators and developers of future serious games:

1. A pedagogically sound serious game concept
2. A solid construct of explicit and implicit learning events that are directly linked to instructional objectives.
3. Purposeful use of gaming technologies to ensure appropriate levels of realism and immersion in relation to the learning content.
4. A robust storyline that is contextually rich and engaging.
5. The so-called “fun factor” which represents an appropriate blend of learner engagement, competition, humor, and entertainment elements.

In addition to these design considerations, when crafting serious games for mental health, it is important to remember the “*Primum non nocere*” principle, also known as “*Do no harm*” and exercise a proper degree of scientific caution.

5 Conclusions

During the course of this effort, the research team uncovered a number of research gaps and practice-related inconsistencies within the field of psychological health education. While they do not necessarily pose a significant threat for the design and development of novel training solutions, it is important to monitor the trends surrounding these issues and structure the future efforts to address the existing set of challenges. Novel psychological health education tools are critically needed to mitigate the increasing psychological healthcare needs of military and civilian populations. While the cited training examples can certainly be considered a step in the right direction, a variety of PH instructional programs are needed for today’s military community. To ensure the effectiveness of the new programs, it is important to leverage the theoretical foundations in the area of deployment psychology, emotional intelligence, and stress resilience, specifically with the emphasis on decision making.

References

1. Castro, C.A., Hoge, C.W., Cox, A.L.: Battlemind training: Building soldier resiliency. In: Proceedings of 1ATO Research & Technology Organization Meeting, Human Dimensions in Military Operations: Military Leaders’ Strategies for Addressing Stress and Psychological Support, France, RTO-MP-HFM-134, Paper 42, pp. 42-1 – 42-6 (2006)
2. Feldner, T.M., Monson, C.M., Friedman, M.J.: A Critical Analysis of Approaches to Targeted PTSD Prevention: Current Status and Theoretically Derived Future Directions. Behavior Modification 31(1), 80–116 (2007)
3. Cannon-Bowers, J., Bowers, C.: Learning and Technology-Based Solutions for PTSD Prevention: An Example of Future Medical Simulation. Paper presented at the Interservice/Industry Training, Simulation and Education Conference (IITSEC), Orlando, FL (2007)
4. Stetz, M., Long, C., Wiederhold, B.K., Turner, D.: Combat Scenarios and Relaxation Training to Harden Medics Against Stress. Journal of CyberTherapy and Rehabilitation 1(3), 239–246 (2008)
5. Wexler, S., Corti, K., Derryberry, A., Quinn, C., Van Barnveld, A.: Immersive Learning Simulations: A 360-Degree Report, eLearning Guild (2008)

6. Andrews, A., Joyce, R., Bowers, C.: Using Serious Games for Mental Health Education. In: Cannon-Bowers, J., Bowers, C. (eds.) Serious Game Design and Development: Technologies for Training and Learning. Information Science Reference, Hershey (2010)
7. Parsons, T.D., Rizzo, A.A.: Affecive Outcomes of Virtual Reality Exposure Therapy for Anxiety and Specific Phobias: A Meta-Analysis. Journal of Behavior Therapy and Experimental Psychiatry 39, 250–261 (2008)
8. Rizzo, A., Reger, G., Gahm, G., Difide, J., Rothbaum, B.: Virtual Reality Exposure Therapy for Combat-Related PTSD. In: Post-Traumatic Stress Disorder Basic Science and Clinical Practice. Humana Press, Totowa (2009)
9. Morie, J.: Re-Entry: Online Virtual World as a Healing Space for Veterans. Institute for Creative Technologies, University of Southern California (2009)

Mixed Reality as a Means to Strengthen Post-stroke Rehabilitation

Ines Di Loreto[1], Liesjet Van Dokkum[2], Abdelkader Gouaich[1], and Isabelle Laffont[2]

[1] LIRMM-Université Montpellier 2
{diloreto,gouaich}@lirmm.fr
[2] Movement to Health Laboratory-Université Montpellier 1
elisabeth.van-dokkum@univ-montp1.fr, i.laffont@wanadoo.fr

Abstract. Purpose of this paper is to present a mixed reality system (MRS) for rehabilitation of the upper limb after stroke. Aim of the system is to increase the amount of training using the fun factor as driver. While the acceptance of such a system can be assessed with patients, true clinical validity can be assessed only through a long randomized clinical trial. However, a first important impression of usefulness can be based on therapists' expertise. For this reason before testing the MRS with patients we carried a study with therapists involving the rehabilitation staff of a French hospital. Three sessions, one using the Wii System with a commercial game, another using an ad hoc developed game on a PC, and another using a mixed reality version of the same game were held. In synthesis results have shown the MR system is regarded to be useful for a larger number of patients, in particular the ones in the more acute phase after stroke.

Keywords: Mixed reality, Post stroke rehabilitation, Serious Games.

1 Introduction

Each year in the United States alone approximately 795.000 people are affected by stroke [3]. In France there are around 150.000 strokes each year [17], and similar trends are evident all around the industrialized world. Being the third cause of mortality and the first one of handicap in adults, we can say that stroke is one of the main health problems in our era [20]. A stroke can cause significant disabilities, such as paralysis, language difficulties, and cognitive impairment. These disabilities strongly affect the quality of daily life for survivors both from the behavioral and physical point of view. Concerning this latter point, several researchers on the field have shown that important variables in relearning motor skills and in changing the underlying neural architecture after a stroke are the quantity, duration, and intensity of training sessions (for more information see e.g. [10]).

In this paper we propose a mixed reality system (MRS) for rehabilitation of the upper limb after stroke. Aim of the system is to increase intensity and number of training session using the fun factor as driver. However while the fun factor can be assessed with patients, clinical validity can be assessed only through long clinical training sessions and/or based on therapists expertise. So, before testing the MRS with patients, we carried a study with therapists in order to understand the potential and the

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 11–19, 2011.

benefits of mixed reality for rehabilitation purposes. The study involved the rehabilitation staff of a French hospital. Three sessions, one using the Wii System with a commercial game, another using an ad hoc developed game on a PC, and another using a mixed reality version of the same game were held.

Before describing the pilot study in detail in the following subsections we will describe in more depth what a stroke is and what its consequences are, and why a 'virtual' approach could be useful for this kind of rehabilitation.

2 Describing Stroke

A Cerebral Vascular Accident (CVA) or simply ‘stroke’ is traditionally defined by the World Health Organization as “a neurological deficit of cerebrovascular cause that persists beyond 24 hours or is interrupted by death within 24 hours” [20]. A stroke induces a loss of brain functions due to disturbance in blood supply to the brain, resulting in altered sensori-motor and/or cognitive functioning. The most striking motor impairment is the loss of motion control on the contra-lesional side. Cognitively one can think of language disturbance, attention troubles and hemi neglect, whereby the contra-lesional side of the body and of the environment is completely neglected.

Because of this, stroke sufferers are often unable to independently perform day-to-day activities such as bathing, dressing and eating, inducing a loss of autonomy and a decrease in quality of life [20]. Stroke is a highly heterogeneous syndrome. The specific abilities that will be affected by stroke depend on the location, type and size of the lesion. Each patient is characterized by a specific combination of deficits [11]. Therefore, stroke rehabilitation programs are strongly personalized and not generic. They are adapted to a particular patient, to regain as much function as possible.

The most important gain of function takes place within the first six months after stroke. In most countries, due to high costs, only 6/8 weeks of rehabilitation are executed under the direct supervision of an expert (i.e. in the hospital or rehabilitation centre). In addition, therapy focusing on mobility (recovery of leg function) is the primary concern within these first weeks, so that recovery of arm function is conducted mainly at home [19].

Important in relearning motor skills and in changing the underlying neural architecture are the quantity, duration and intensity of training sessions (see for example [14][9]). To experience significant recovery, stroke patients must then perform a substantial number of daily exercises. Unfortunately, typical rehabilitation sessions with therapists include a relatively small number of motions [8] due to the low duration of personal rehabilitation sessions. Home training can increase the amount of executed movements. However, while this is often prescribed, only 31% of the patients actually perform the recommended exercises [16]. This raises the problem of selecting a home-based rehabilitation technology that both can help and motivate patients to perform their therapeutic exercises to ameliorate their recovery after stroke.

3 Motivations: Advantages of 'Virtual Rehabilitation'

The challenge for post stroke rehabilitation is then to create exercises able to decrease the monotony of hundreds of repeated motions. In order to overcome this challenge, different kinds of 'non traditional' therapies have been proposed.

For example, the possibility of using 'virtual' rehabilitation has been the subject of experiments by several authors (such as [13][7]). Although most of the studies on this topic are linked to the study of virtual reality environments recent works have focused on the use of videogames and consoles for rehabilitation (such as [2][5]).

We can summarize the results of these studies as follows:

1. *Personalization*: Virtual rehabilitation technologies create an environment in which the intensity of feedback and training can be manipulated to create the most appropriate, individualized motor learning paradigm [7].
2. *Interactivity:* Virtual rehabilitation exercises can be made to be engaging so that the patient feels immersed in the virtual world. This is extremely important in terms of patient motivation [12], which in turn, is one of the key factors for recovery.
3. *Feedback:* Interactive feedback can contribute to motivation. By providing visual and auditory rewards such as displaying gratifying messages in real time, patients are motivated to exercise [4][6].
4. *Tracking:* The evolution of the patient's performances can be easily stored, analysed and accessed by therapists [7][12].
5. *Telerehabilitation*: Virtual rehabilitation can stimulate the patient using a variety of rehabilitation exercises at a low cost. In addition cheaper personal equipment (pc-based for example) could eventually allow rehabilitation stations to be placed in locations other than the rehabilitation centre, such as patient's home.

On the other hand, Virtual Rehabilitation does raise significant challenges before its widespread adoption such as:

- Clinical acceptance, which relies on proved medical efficacy.
- Therapist's attitude towards the technology (e.g. the therapist fears that technology could replace therapists and the like).
- Patient's attitude towards the technology (e.g. the patient may not consider a game to be 'real' rehabilitation).
- Ethical and regulation challenges linked to the kind of technology used.

In the rest of this paper we will describe: (i) the mixed reality system and the game we designed for the experiment, (ii) the experiment held and its results, (iii) some consideration on future improvements on the system.

4 The Mixed Reality System

In this section we describe a mixed reality system (MRS) conceived for post stroke rehabilitation purposes. In general a mixed reality environment aims to merge real and virtual worlds to create a new context of interaction where both physical and digital objects co-exist and interact consistently in real time. The mixed reality system we conceived and developed for this study is composed of three main subsystems: (i) the gaming subsystem responsible for managing the game application; (ii) the motion capture subsystem responsible for tracking patient's movement; and (iii) the display subsystem responsible for displaying the virtual game environment in a physical environment.

The gaming subsystem follows a client-server pattern: the mechanics of the game are implemented at the server level. The game server translates this input into game events that are used to compute the next state of the game according to the game mechanics. Once the new state of the game has been computed, update events are sent to the game client(s) to update their local copies of game objects, which in turn update the graphical environment. The game client fulfils the following functions: (i) receiving patient's movement events from the motion capture system; (ii) forwarding these events to the game server; (iii) receiving events from the game server to update game object state; and finally (iv) generating the game's output in terms of 3D scenes and sounds.

The motion capture subsystem is responsible for tracking the patient's hand and to communicate these movements to the game client. The motion capture system is composed of three components:

1) An IR emitter: this is an active sensor built using a simple infrared (IR) emitter that is attached on the patient's hand using a Velcro strap. This basic device is very cheap (less than 1 euro) and convenient to use even with patients that suffer from spasticity.
2) An IR camera: To track the IR emitter a Nintendo Wiimote was used as an IR camera. The Nintendo Wiimote is an affordable device (less than 30 euros) that has been used in many open source projects. Consequently, a plethora of open source libraries now offer the use of Wiimote as pointing device.
3) A processing software: the role of this component is to translate Wiimote events into mouse events so that the IR emitter and IR camera are considered as a standard pointing device from the operating system's point of view.

T*he display subsystem:* Within a mixed reality context it is necessary to display game objects directly in the physical world. Thanks to a new generation of mini video projectors, namely pico projectors, it is possible to display the game on a flat surface almost anywhere. Within the context of this experiment, the pico projector is attached to a support to display the game onto a standard table. Pico projectors are affordable (less than 200 euros), lightweight (less that 300g), and easy to use by simply plugging a VGA cable into a computer. However, they are limited in terms of brightness (the one that was used in this experiment was less than 30 lumens). This constraint has to be taken into account when building the game's graphical environment by selecting colours and contrasts to compensate for this limitation.

The overall system is presented in Figure 1. The patient holds on his hand the IR emitter device and sits in front of the table (a) to play the exercise. The IR camera (c) and pico projector (d) are placed on the top of a support (b). A laptop computer is used to run the game client and to establish a Bluetooth connection with the WiiMote.

The ad hoc developed games: The MR system described above is able to support different types of therapeutic games, conceived ad hoc to train different types of movements. A module of the game server allows developers to integrate different types of adaptation for every game, offering a personalized gaming experience for each patient.

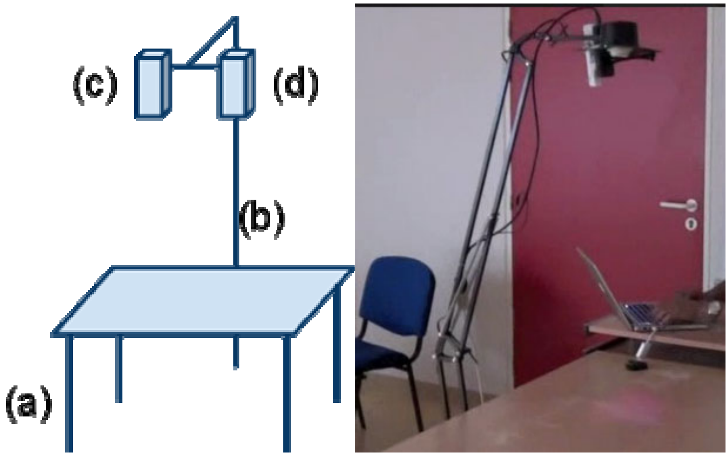


Fig. 1. The overall system

5 Therapists Opinion and Virtual Rehabilitation Systems

A great majority of systems created for virtual rehabilitation purposes tests patient's acceptance and do not test in a systematic way therapists' opinions about the system (interesting exceptions are [1], [18]). On the contrary we are convinced that only a therapist can assess the rehabilitation value of a system. In fact, from a previous pilot study we held with 4 patients affected by stroke emerged clearly that the patient is not able to measure the impact of the system on her/his rehabilitation session/program. In that pilot study patients tested a game with a graphical tablet, a mouse and the mixed reality system described above. While we will not enter in detail here, we can summarize results as follow. Firstly patients were not able to differentiate the difficulty of using the system from the difficulty of executing the movement, strongly impaired by the stroke. Secondly, the clinical eye of the therapist, who assisted the sessions, was able to observe for which patient the exercise was more difficult or when the patient had to put all efforts to move his hand (and thus was concentrating on his hand and not on the screen and the like). For this reason we decided that before testing the system with patients, it is really important to assess potential benefits with the therapists.

5.1 The Experiment Protocol

As said in the introduction aim of the experiment described in this paper was to compare the use of an in-house designed Mixed Reality System (MRS) with two alternative tools through an empirical investigation. The two alternative tools were a classical PC system, and the Wii system. In order to compare acceptance level of these systems (in Shackel terms, see [15]) we decided to assess perceived utility, usability, likeability (affective evaluation) of the three systems.

Participants: The mixed reality system described above has been tested on 17 members of the medical staff of a rehabilitation department of a French hospital. 3 occupational therapists, 9 physiotherapist, 3 students in physiotherapy and 2 general practitioners participated in the study. Participants had an average of 15 years of experience.

The experiment: The experiment was composed of different sessions, each one testing a different 'form of potential therapy' (the Wii console with the Wiimote, computer and mouse, mixed reality with captor). The game on the computer and the one on the mixed reality system were identical. This was done in order to show that games and hardware system could be independent. The virtual rehabilitation game developed for this experiment was conceived to train upper limb 2D movements (i.e. movements on a plan without need to lift the arm). The gameplay stages a maize plantation and crows trying to steal maize. Aim of the game is to chase away all the crows; aim of the exercise is to train free movements. For each session therapists were let free to play with the system, but they were asked to keep in mind that they were playing the patient's role. Each session took 10 minutes. After the last session a questionnaire was conducted. Order of sessions was randomized.

The Questionnaire: The survey was conceptually divided into two parts. Results are expressed as a note on a four level scale, "4" being the best note and "1" being the lower one (Likert scale). The first part addressed the different hardware systems while the second one the games. In the first part, several questions addressed the usability of the different systems from the patient's point of view. Another group of questions addressed the perceived utility of the system for therapy. Finally, a last question compared usability of the three devices. In the second part of the survey, a set of questions addressed the different games utility for therapy, and the fun perception of each game. Each question asked also to motivate the rating.

6 Results and Discussion

As said previously, the first set of questions was related to the system and did not concern the games.

The first question asked the therapists if they thought that patients **will be able** (from a physical point of view) to work with the three systems (considered as a whole). MR rated a means of 3,8(±0,4); PC rated 3,1(±0,9), the Wii system rated 2,9(±0,9). Another question addressed the **devices** used with the three systems (the Wiimote, the mouse, the led captor). The u**sability** for a patient of the IR Led rated a means of 3,3(±1,2), the mouse 2,2(±1,2), the Wiimote 2,4(±1,4).

To motivate these ratings the therapists wrote in synthesis that to be manipulated the Wiimote requires a set of motor skills that patients do not necessarily have regained. Most patients have also coordination problems. For those patients it will be difficult to support or release buttons on the controller in time. In addition, having to press a lever can generate spasticity or synkinesia (i.e. uncontrolled movements). While the subject of spasticity was raised by all the therapists (for both, the mouse and the Wiimote) several of them underlined that the Wiimote could be interesting in order to re-educate grasping.

On the other hand, in the therapists' words, the MR system is easy to understand and requires no cognitive effort by the patients. Finally, the sensor on the hand seems to me more suited for a larger number of patients.

Almost all the questionnaire in the end commented that the two systems (MR and Wii) could complement each other following the patient recovery. While the mixed reality system can be used in a more acute phase, the Wiimote becomes an interesting device for patients who have recovered better (remembering that not all the patients will attain this phase).

The therapists were then asked if they thought that working with the different systems could be **useful** for the patients. The MR rated 3,8(±0,4), The PC rated 3,1(±1,1) and the Wii rated 3,5(±0,6).

As for the usability questions, both, the Wii and the MR have their advantages. The mixed reality system allows for coordination and proprioception (i.e. the sense of the relative position of neighbouring parts of the body). In fact, the patient must try to move her/his limb in coordination with other parts of his/her body to reach a target. The MR allows also for a better hand-eye coordination, as the hand moves within the game. In fact in both, the PC case and the Wii case, the patients have to look at another place (the screen) in order to perform the task. This kind of task requires an additional difficulty for patients in the first rehabilitation phase. Comments on PC usage are also related to limitations about screen dimension. On the other hand the Wii uses more planes in the space respect to our MR version which is more useful for joints. Finally, both allow the therapist to work with larger amplitudes than the traditional PC.

The second set of questions was related to games. The first question addressed the **usefulness of the** ad-hoc **game** with respect to the commercial Wii game. As perceived utility the ad hoc game rated 3,6(±0,6) while the Wii game rated 2,9 (±1,0). On the other hand the perceived fun of the ad hoc game was 3,0(±1,0) while the Wii game rated 3,6 (±0,5). This is an indicator of two elements. While there is a need to create ad-hoc games for stroke rehabilitation, they need to be real games with the fun factor and not only simply exercises. Also in this case there are different needs in different parts of the rehabilitation. In fact, more complex games require patients with no cognitive or attention difficulties.

The final question comparing the utility for therapy of the three systems had the following results: MR rated 3,6(±0,7), PC 2,9 (±1,1) and the Wii 3,2(±0,7).

In this question one of the therapists stated clearly that the MR system is more suitable for patients in rehabilitation even if commercial games seem more fun.

To summarize, the Mixed Reality System has a higher potential to be useful for a huge number of patients. In fact, the most underlined problem with PC and Wii for rehabilitation - the use of a hand held device - becomes one of potentiality of mixed reality. In addition, the mixed reality system uses objects from the natural environment as an interface to the computer. This intuitive way of interaction simplifies the communication between man and machine, especially for unpractised users. Finally, a MR system can be used by several persons because it is not restricted to one screen or one keyboard.

7 Conclusion and Future Works

Objective of this paper was to evaluate a mixed reality system (MRS) for rehabilitation of the upper limb after stroke. Because it is the goal of therapists to guide the stroke patients through their rehabilitation, we are convinced that the therapists are the most suited to understand the potential utility of a tool for rehabilitation.

For this reason a study has been conducted in a rehabilitation department of a French hospital in order to understand the potential and benefits of mixed reality. The study involved physiotherapists and occupational therapists of the same hospital. Three sessions, one using the Wii System with a commercial game, another using an ad hoc developed game on a PC, and another using a mixed reality version of the same game were held. In synthesis results have shown the MRS is seen as useful for a huger number of patients, in particular the ones with more acute diseases. Therapists found value also in the Wiimote for patients with a good recovery to exercise grasping while they found no interest in the use of a PC system.

Remarks on the MR and the Wii systems underlined the necessity of objects manipulation to recover grasping. However this is a limitation of the above described prototype not a mixed reality limitation.

While the MR presented in this paper did not use objects and was conceived to work in a 2D space we are currently working on another prototype adding physical objects manipulation. This second prototype will also take into account a user profile using agents to create a personalized rehabilitation experience. In fact the most innovative and interesting part of this kind of rehabilitation tools is the mix of adaptive aspect combined with the ease of use of the same tool. Further prototypes will extend objects manipulation in the 3D space. This way the interaction with virtual objects can be done in the same way as the interaction with real objects without removing the personalization part. The adaptation algorithm can in fact place targets in the best place to push the patients to work exactly the area in need. In addition both hands can be used for interaction because the user does not have to hold a device. Several objects could be then being manipulated at once enabling bimanual therapy.

Finally, the last prototype we will develop will implement collaboration between several users in order to understand if a collaborative therapy is possible.

Acknowledgements. This work was performed as part of the MoJOS project (http://www.mojos.fr/home/). The authors wish to express their gratitude to the Gui de Chauliac Hospital staff for their cooperation and to Philippe Armingaud for his valuable contributions.

References

1. Annema, J., Verstraete, M., Abeele, V., Desmet, S., Geerts, D.: Videogames in therapy: a therapist's perspective. In: Proceedings of the 3rd International Conference on Fun and Games (Fun and Games 2010), pp. 94–98. ACM, New York (2010)
2. Burke, J.W., McNeill, M.D.J., Charles, D.K., Morrow, P.J., Crosbie, J.H.: McDonough, S. M.: Optimising engagement for stroke rehabilitation using serious games. The Visual Computer 25(12), 1085–1099 (2009)

3. Center for Disease Control and Prevention: Stroke facts (2010), http://www.cdc.gov/Stroke/index.htm
4. Chen, Y., Huang, H., Xu, W., Wallis, R.I., Sundaram, H., Rikakis, T., Ingalls, T., Olson, L., He, J.: The design of a real-time, multimodal biofeedback system for stroke patient rehabilitation. In: Proceedings of the 14th annual ACM international conference on Multimedia - MULTIMEDIA 2006, p. 763 (2006)
5. Flynn, Sheryl, Palma, Phyllis, Bender, Anneke: Feasibility of Using the Sony PlayStation 2 Gaming Platform for an Individual Poststroke: A Case Report. Journal of Neurologic Physical Therapy 31(4), 180–189 (2007)
6. Goude, D., Bjork, S., Rydmark, M.: Game design in virtual reality systems for stroke rehabilitation. Studies in health technology and informatics 125(4), 146–148 (2007)
7. Jack, D., Boian, R., Merians, A., Tremaine, M., Burdea, G., Adamovich, S., Recce, M., Poizner, H.: Virtual reality-enhanced stroke rehabilitation. IEEE Transactions on Neural Systems and Rehabilitation Engineering: a Publication of the IEEE Engineering in Medicine and Biology Society 9(3), 308–318 (2001)
8. Lang, C.E., Macdonald, J., Gnip, C.: Counting repetitions: an observational study of outpatient therapy for people with hemiparesis post-stroke. Journal Neuro. Phys. Therapy 1(31), 3–10 (2007)
9. Langhorne, P., Coupar, F., Pollock, A.: Motor recovery after stroke: a systematic review. The Lancet Neurology 8(8), 741–754 (2009)
10. Micera, S., Carrozza, M.C., Guglielmelli, E., Cappiello, G., Zaccone, F., Freschi, C., Colombo, R., Mazzone, A., Delconte, C., Pisano, F., Minuco, G., Dario, P.: A Simple Robotic System for Neurorehabilitation. Autonomous Robots 19(3), 271–284 (2005)
11. National Stroke Association. Effects of Stroke (2010), http://www.stroke.org/site/PageServer?pagename=EFFECT
12. Popescu, V.G., Burdea, G.C., Girone, M., Bouzit, M., Hentz, V.R.: Orthopedic Telerehabilitation with Virtual Force Feedback. IEEE Transactions on Information Technology in Biomedicine 4(1), 45–51 (2000)
13. Rizzo, K.: A swot analysis of the field of virtual reality rehabilitation and therapy. Presence: Teleoperators and Virtual Environments 14(2), 119–146 (2005)
14. Selzer, M., Clarke, S., Cohen, L., Duncan, P., Gage, F.: Textbook of Neural Repair and Rehabilitation. Medical Neurorehabilitation, vol. 2. Cambridge Uni. Press, Cambridge (2006)
15. Shackel, B.: Usability – Context, framework, definition, design and evaluation. In: Shackel, B., Richardson, S.J. (eds.) Human Factors for Informatics Usability, pp. 21–31. Cambridge University Press, Cambridge (1991)
16. Shaughnessy, M., Resnick, B., Macko, R.: Testing a model of post-stroke exercise behavior. Rehabilitation Nursing: The Official Journal of the Assoc. of Rehabil. Nurses. 1(31), 15–21 (2006)
17. Societe Francaise Neuro-Vasculaire, bilan (2010), http://sfnvfrance.com/
18. Tam, S.: An Environment for Stroke Therapy Game Authoring. PhD Thesis. Department of Computer Science & Engineering - Washington University in St. Louis (2010)
19. van der Lee, J.H., Wagenaar, R.C., Lankhorst, G.J., Vogelaar, T.W., Deville, W.L., Bouter, L.M.: Forced Use of the Upper Extremity in Chronic Stroke Patients: Results From a Single-Blind Randomized Clinical Trial. Stroke 30(11), 2369–2375 (1999)
20. World Health Organisation, http://www.who.int/en/

A Virtual Experiment Platform for Mechanism Motion Cognitive Learning

Xiumin Fan, Xi Zhang, Huangchong Cheng, Yanjun Ma, and Qichang He

CIM Research Institute, Shanghai Jiaotong University, Shanghai, 200240, China
xmfan@sjtu.edu.cn

Abstract. In order to give students a more intuitionistic understanding in mechanism motion system, a virtual experiment platform is designed and developed. First, experimental component models, which contain both visual information and logical information, are built. The logical information is modeled according to a Unified Modeling Language called Modelica (MO). Then, a virtual experiment scene, which is described by Modelica, is assembled in a virtual reality environment. The virtual scene MO model is flatted to a set of equations which are compiled and solved, and the mechanism motion data can be output. Last, the motion data are exported into the Virtual Reality environment for the simulation result visualization. Students can use the platform to build mechanism system experiments and simulate the component motion for a better understanding of mechanism composition and its movement principle. The platform is universal and can be expanded to other subjects easily because the experimental components are built by Unified Modeling Method.

Keywords: virtual experiment, mechanism motion experiment, modelica modeling, virtual reality.

1 Introduction

In the traditional experiment for university teachings, a hardware experiment environment is required for students to do experiment. It is used to taking up a lot of teaching resources. In recent years, the increase of the student quantity brings a great pressure to the experiment teaching. A virtual experiment teaching platform can relieve the stress and provide approach to improve the innovation ability of students.

In the past 30 years, different kinds of object-oriented simulation tools were developed with the progress of computer technology, such as PSpice for circuit simulation, Adams for multi-body system simulation, Modelica (MO) for general multi-domain modeling and simulation. Modelica Language is proposed as an object-oriented, equation-based and unified modeling language to solve complex multi-domain physical system simulation [1][2]. In China, the research of virtual experiment for university students is in an initial stage. Some universities have obtained some results, such as, the college physics 3D online virtual experiment based on Virtools in Jilin University, the virtual experiment platform base on Jini technology in Sichuan Normal University, the remote virtual laboratory based on LabView in Taiyuan University of Technology, the virtual laboratory for intelligent

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 20–29, 2011.

control based on VRML and Matlab in Zhengzhou University [3]. While in other countries, the research of virtual experiment was started earlier and many results have been achieved, such as virtual laboratory for information security teaching in Columbia State University, multimedia virtual laboratory in Swiss Federal Institute of Technology, and e-Learning virtual laboratory in Purdue University [4].

However, there is no general development platform for virtual experiment, most current virtual experiments are developed for specific subjects. This single subject development mode limits the dissemination of virtual experiment in education. In this paper, according to the consistency of the intrinsic description in mathematical equation about the physical law of different subjects [5], the object-oriented modeling language Modelica is adopted and used to build the unified models of experimental components for different subjects. Based on these experiment component models, the relationships among components are connected, and the experiment scene model is constructed, and then the experiment model is solved to get results. By this approach, a general virtual experiment platform can be constructed so as to support different virtual experiments for multi-subjects.

Based on this unified modeling and equation solving idea, a virtual experiment platform for mechanical principle course is built to support the cognitive learning of mechanism motion experiment. The organization of the following contents is: Section 2 introduces the general idea and system structure; Section 3 introduces the key methods and techniques; Section 4 introduces the development and application of the system; Section 5 summaries all the contents.

2 General Idea and System Structure for Virtual Experiment Platform

The system structure is shown in Fig.1. Firstly, basic components are abstracted from the mechanism motion components by analyzing the logical relationship among the various components of mechanism motion system experiments for mechanical principles course. The logical model of the basic component is modeled by Unified Modeling Language (Modelica). This logic model together with component's 3D model is integrated to form a completed component information model which is described by XML. The entire XML component models can be formed as a component list. Then, users can select necessary components from the component list and connect these components according to their interface types, so as to build a virtual model of the experiment and save the model into a MO file. The experimental scene files are compiled and resolved with the support of the logical MO model. Next, with the support of component's logic MO model, a solver is executed to compile and calculate the virtual experiment MO model to get calculation results. Finally, the calculation results are analyzed and useful data about the mechanical component geometric motion are extracted. These geometric motion data file is used to drive the components' movement in 3D visualization environment. The key technologies include modeling of component logical information, building and solving the experiment scene model and visualization of the simulation results. The virtual environment for virtual experiment assembly and simulation is developed based on a virtual reality application platform called "VR Flier" [6-7].

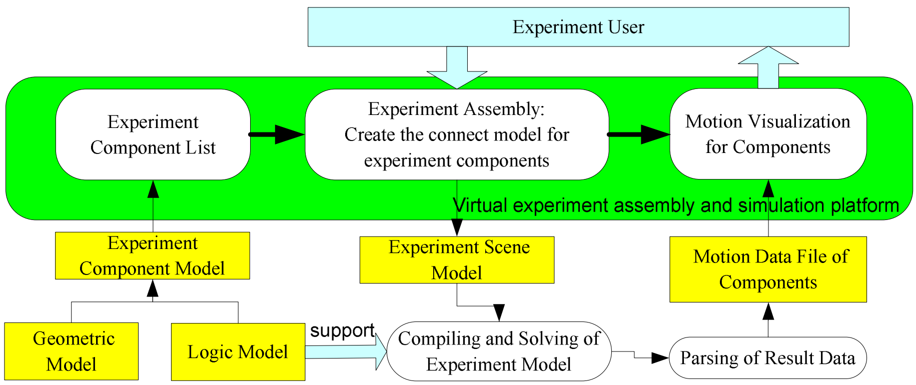


Fig. 1. System structure for virtual experiment platform

3 Method and Technology

3.1 Component Modeling and Information Mapping

Component Logic Information Modeling with Modelica. In order to build logic model of typical experiments in mechanical kinematics course, basic components that being abstracted out of these experime1wnts should be unified modeled. These basic components including bar, rotational interface, translational interface, power source, fixed translation and signal interface. These basic components are modeled in Modelica language (shown in Table 1).

Table 1. Logic models of basic components

Name of basic components	Function	Property
Bar	Define the length between two points	Two endpoints' coordinate : *x, y*; Bar length: *l*; angle between bar and x axis: *phi*
Translational interface	Constrain the relative translation motion between two components	Two endpoints' coordinate : x, y
Rotational interface	Constrain the relative rotation motion between two components	Two endpoints' coordinate :x, y
Fixed translation	Define the fixed displacement between two points	Two endpoints' coordinate :x, y
Power source	Drive a component to rotate or translate	Motion parameter output signal
Signal interface	Transfer non-motional signals between two components	Motion parameter output signal

Typical experiment components are combination of these basic components, taking the simple punching machine experiment as an example, its typical experiment components list is shown in Table 2. The numeric character for each table cell means how many numbers of basic components are included for each experiment component.

Table 2. Experiment component list for Simple Punch Machine

Basic Component / Experiment Component	Bar	translational interface	rotational interface	fixed translation	power source	signal interface
Base	0	1	2	3	0	0
Flywheel	1	0	2	0	0	1
Slider	1	1	1	0	0	0
Driver	0	0	0	0	1	1
Punch	1	1	1	0	0	0
connecting bar	1	0	2	0	0	0
Plug	3	1	2	0	0	0

Complete Component Model Description in XML Model. An experiment component's complete information, which includes component's three-dimensional visualization information and logic information from Modelica model, is described in XML (Extensible Markup Language) format. The component XML model contains five types of information as shown in follows:

(1) Component's basic information

The basic information includes the component's ID, Name, Type, Subject information, Creator, Creation date. Each component in MO library has its unique ID.

(2) Component's interface list

The interface contains following information as shown in XML format.

<interface>
<name></name> //Interface name
<type></type>//Type declared in Modelica
<MotionType></MotionType>//Motion style in platform (translation, rotation or fixed)
<Position-x></Position-x>//Interface coordinate in x axis
<Position-y></Position-y>// Interface coordinate in y axis
<Position-z></Position-z>// Interface coordinate in z axis
<Normal-x></Normal-x>//Interface normal vector
<Normal-y></Normal-y>
<Normal-z></Normal-z>
</interface>

(3) Component's property list

The property information includes property ID, Name, Symbol, Value, Unit, Max Value, Min Value, Display Flag, Modifiable Flag and Coordinate Flag. The Display Flag defines whether the result of the property should be returned and displayed in the virtual environment. The Modifiable Flag defines whether the value of the property could be modified by user. The Coordinate Flag defines whether the parameter is related to the three-dimensional model's coordinate system.

(4) Component's logical location in Modelica library

The Modelica library contains all of components' MO logic models which are organized in a tree structure. A MO library named VElibrary has been created. A Component's MO logical location should be included in the XML description. The logical location designates MO model's full path from root node to child node, such as Modelica.VElibrary.Mechanism.MechanismAdditions.Basic.VirtualRotor.

(5) Component's geometrical information

All of Component's 3D geometric models are also stored in a model library. Component's three-dimensional models are in flt format. The same as a component's MO model, a Component's 3D model logical location should be included in the XML description, too. As the geometric modeling coordinate system may be inconsistent with the virtual environment coordinate system, a coordinate transformation matrix should also been included.

Information Mapping from Modelica Model to XML model. In the virtual experiment environment, component information models are stored in XML model format. It is necessary to map component's information from MO model to XML model. The MO model includes two parts, component declaration and function definition. As the XML model is not involved in solving process, it only needs the information of component declaration. When the solver is called and used to calculate an Experiment Scene Model, the solver will call the component's MO model which is corresponding to its XML model. The information mapping process includes five steps:

(1) Traverse the interface definition file in Modelica library to generate a **connector** class list. All of the **connector** interface components can only be called by **reference**.
(2) Traverse the interface definition file in Modelica library to generate a **partial model** class list. All the **partial model** can only be called by **extends**. The **partial model** may package several **connector** models to simplify a component's structure.
(3) A MO file may include one MO component or several MO components. Its logical hierarchy can be constructed by prefix **package**, **model**, **record**, **type**, **connector**, **block** and so on [8]. When mapping a MO file, a Model List is generated to record the components defined in the MO file. During the **Model List** generation procedure, **extends** and **reference** members in each component are traversed. **Extends** members are searched in **partial model** list and **reference** members are searched in **connector** list to get the interface list for each component.
(4) Generate a flat model of each component. A flatten algorithm [9] is adopted to get a flat model. The flat model includes parameter information and the mathematical relationship among parameters. The parameter information, including symbol, initial value, maximal value, minimal value, unit [10], modifiable flag, display flag and coordinate flag, can be extracted from the parameter declaration in flat model. For each parameter, only symbol is indispensable.
(5) After above four steps, most of the component's information in XML model is acquired. However, the parameter's name should be added by a modeler, as the name of parameter's meaning is not included in a MO model, and user can distinguish the meaning of each parameter clearly.

3.2 Virtual Experiment Scene Assembling and Calculation

The virtual experiment platform consists of two main modules: experiment assembling module and solving module.

For experiment assembling, user selects components (described as XML file) according to a certain experiment and connects the interfaces of the components (the interfaces can be connected only when the type of two interfaces is identical). For example, in mechanism kinematic motion experiments, the relative motion interface between components consists of three types: translation, rotation and logical connection; components having the same motion interface can be connected together. User can modify the parameter value of a component if its parameter is modifiable. After the experiment is assembled, user sets the simulation parameter. A MO file describing the virtual experiment scene is generated for compiling and calculation.

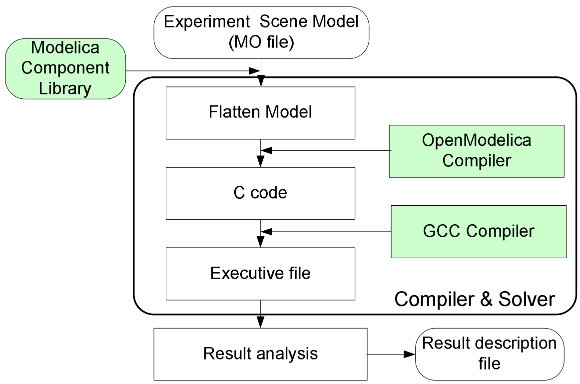


Fig. 2. Compiling and solving process for experiment scene

The compiling and solving process is as shown in Fig.2. With the experiment scene MO file ready, the Compiler & Solver load in the component's Modelica library beforehand, and then, compile and calculate the experiment scene MO file according to the simulation time setting. Final, a result description file is returned back to the virtual experiment environment. The compiling and solving process include several steps. Firstly, the experiment scene MO file is transferred to flat model. Then, the OpenModelica[11] compiler transfers the flat model into C code. Next, the GCC compiler compiles the C code into executive file. Finally, the executive file is used for result calculation.

3.3 Visualization of Calculation Results

After the experiment scene model is compiled and solved, the geometrical movement data of each moving component can be output, the data is in PLT format. In order to visualize the 3D motion of experiment components in VRFlier platform, a data conversion interface is designed for converting PLT format data to the data format specified in VR Flier, so as to drive the component's movement in virtual environment. The conversion process is as following: first, truncate the necessary data from PLT file; second, read the truncated data file into system memory through C++ I/O operation and assign the data to a corresponding array; third, transform the position coordinates and angles into transformational matrixes.

After the data processing, a path file of components' movement can be available. With this path file, a model-load-list file for loading geometrical component models can be created. The order of loading models should be corresponded to the order of the output key points. VR Flier platform can drive the movement of 3D components' models in virtual environment by loading the motion data file, and the time parameter between two animation frames can be set if necessary.

4 Development and Application of the Virtual Experiment Platform for Mechanism Motion Cognitive Learning

4.1 Development of the Virtual Experiment Platform for Mechanism Motion

Based on VR-Flier, a general platform for virtual reality application development, a virtual experiment platform for mechanism motion learning has been developed, as shown in Fig.3.

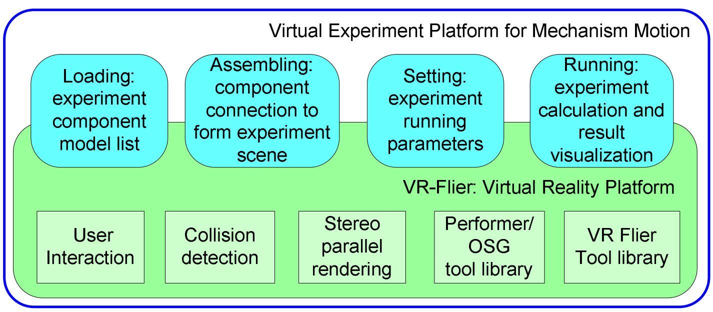


Fig. 3. Structure of software modules of the virtual experiment platform for mechanism motion

Those methods and technologies mentioned in the above section have been embedded into this VE platform. Software modules have been developed, such as "Loading: experiment component model list", "Assembling: component connection to form experiment scene", "Setting: experiment running parameters" and "Running: experiment calculation and result visualization". Some mechanism motion cognitive learning experiments, such as simple punching machine, crank & rocker mechanism and traveling guide rod mechanism, have been tested.

4.2 Assembly and Simulation of Simple Punching Machine Experiment

Taking the simple punching machine as an application example, the virtual experiment process is as follow:

Step 1: Load experiment components into virtual environment, including flywheel, base, slider, connecting bar, plug, driver and punch, as shown in Fig.4.

Step 2: Add a connection between two experiment components which have the same matching interface. For example, select the matching interface of connecting bar and plug to establish a connection, as shown in Fig.5. By this mean, all connecting components can be assembled together to build a simple puncher, as shown in Fig.6.

Step 3: Set parameter values of each component's property. Select the driver component and set its rotational speed, as shown in Fig.7.

Fig. 4. Experiment components of a simple punching machine

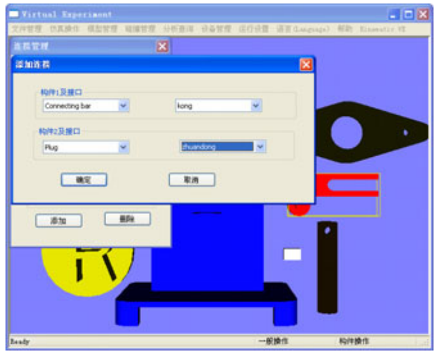


Fig. 5. Add a connection between experiment components

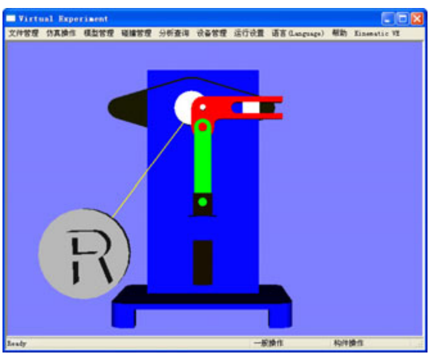


Fig. 6. All of components connected together

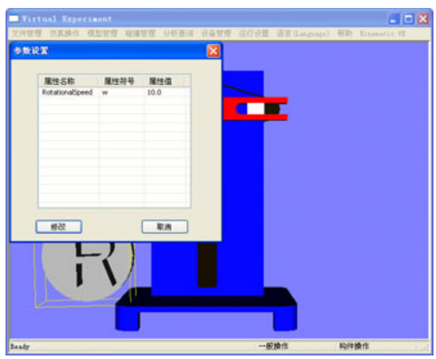


Fig. 7. Set parameters of component properties

Step 4: Set running parameters of simulation, including start time, end time and total number of intervals, as shown in Fig.8.

Step 5: Select "start" button from menu of "virtual experiment of mechanism motion" to run the experiment. The result of experiment simulation can drive the 3D component model to move and achieve the motion of punching, as shown in Fig.9.

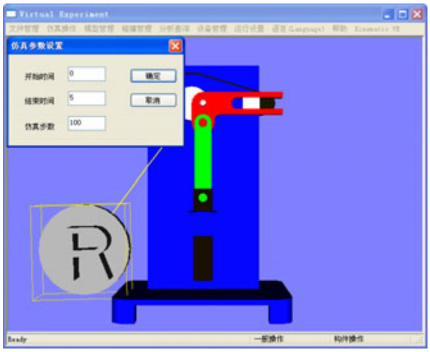


Fig. 8. Set operating parameters of simulation

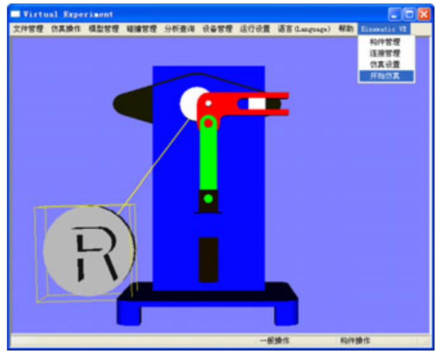


Fig. 9. Start simulation of the virtual experiment

4.3 Discussion

In this virtual environment for mechanism motion experiment, typical experiments can be assembled and simulated; it provides students with a convenient software tool for mechanism motion cognitive learning by a DIY approach. This also means the physical experiment in laboratory can be replaced by virtual experiment to a certain degree.

The development and application of the VE platform provides students with an assembly platform helpful to assemble and simulate experiment models, which enables students to better understand the assembly process of mechanism and the principle of mechanism motion. But this VE platform is still in its first version and has some shortcomings, such as the need of continuous improvement of the realism of virtual scene, and the need of a better and intuitive interaction, and even more functions, so as to let students have a more realistic feeling of doing real experimental environment.

5 Conclusion

In this paper, a virtual experiment platform for mechanism motion learning based on unified modeling method has been proposed. As the principle of mechanism motion can be described by a set of mathematical models, the experiment process can be modeled with a unified method. The method and technology, including experiment component modeling, experiment scene model assembling and solving, experiment result visualization, are discussed in detail. The virtual experiment component model contains two types of information, the geometrical information and the logical information. Basic components of mechanism motion experiment have be abstracted by analyzing the motion mechanism and modeled by Modelica language. The experiment scene MO model can be flatten to a set of equations and compiled and solved. The final experiment output data can be visualization in a 3D virtual environment. Based on these methods and technologies, together with the support of VR-Flier (a general development platform for virtual reality application), a virtual experiment platform has been developed. In this VE platform, students can assemble a mechanism motion experiment and do simulation by themselves, so as to have a better understanding of the composition of mechanism and its motion principle.

In addition, the virtual experiment platform is based on Unified Modeling Method. Three main modules, the component modeling, experiment compiling & solving, and result visualization, are independently. The structure of this VE platform is flexible and easy to be expended to experiments of other subjects, such as mechanical control.

Acknowledgments. The work is supported by the State Science and Technology Support Program of China (Grant No. 2008BAH29B03). The authors are also grateful to the editors and the anonymous reviewers for helpful comments.

References

1. Fritzson, P., Engelson, V.: Modelica- A unified object- oriented language for system modeling and simulation. In: Eric, J. (ed.) Proceedings of the 12th European Conference on Object- Oriented Programming, pp. 67–90. Springer, Brussels (1998)
2. Liu, J., Huang, Y., Chen, L., Wang, Q.: Mult -Body Dynamics Modeling Based on Modelica. China Mechanical Engineering 21(9), 1088–1089 (2010) (in Chinese)
3. Shi, L., Wang, Z., Fei, M.: Design and implementation of an intelligent control virtual laboratory based on VRML and MATLAB. Chinese Journal of Scientific Instrument, 2594–2597 (2008)
4. Averitt, S., Bugaev, M., Peeler, A., Shaffer, H., Sills, E., Stein, S., Thompson, J., Vouk, M.: Virtual Computing Laboratory(VCL). In: Proceedings of the International Conference on the Virtual Computing Initiative, May 7-8, pp. 1–16 (2007)
5. Schamai, W., Fritzson, P., Paredis, C.: Towards Unified System Modeling and Simulation with ModelicaML:Modeling of Executable Behavior Using Graphical Notations. In: Proceedings 7th Modelica Conference, Como, Italy, September 20-22 (2009)
6. Pi, X.-Z., Fan, X.-M., Yan, J.-Q.: VR Flier: A software platform for virtual reality general application development. Chinese Journal of System Simulation 17(5), 1157–1162 (2005) (in Chinese)
7. Hu, Y., Wu, D.-L., Fan, X.-M.: Virtual Reality Development Platform for Parallel Rendering. Journal of System Simulation 21(15), 4714–4721 (2009) (in Chinese)
8. Fritzson, P.: Principles of Object-oriented Modeling and Simulation with Modelica 2.1. IEEE Press, Hoboken (2004)
9. Tiller, M.M.: Introduction to Physical Modeling with Modelica. Kluwer Academic Publishers, London (2001)
10. Modelica WWW Site [EB/OL]. Modelica Group (February 10, 2006), http://www.Modelica.org
11. OpenModelica WWW Site [EB/OL]. OpenModelica (December 30, 2010), http://www.openmodelica.org/

Mechatronic Prototype for Rigid Endoscopy Simulation

Byron Pérez-Gutiérrez, Camilo Ariza-Zambrano, and Juan Camilo Hernández

Virtual Reality Center – Davinci Research Group,
Nueva Granada Military University, Bogotá, Colombia
byron.perez@ieee.org

Abstract. Haptic systems include hardware and software components for providing programmable sensations of mechanical nature, such as those related to the sense of touch. This article covers the mechatronic design of a rigid endonasal endoscopy simulator that allows the user to feel force feedback collisions with the anatomical structures during navigation in a virtual environment. The mechatronic system design provides tactile feedback information with three degrees of freedom to the user based on an open loop implemented control. The tests were performed on a computational prototype that allows the visualization of medical image volumes and navigation with collision detection system.

Keywords: Haptic, Virtual Reality, Virtual Endoscopy.

1 Introduction

Virtual reality systems are important tools when performing various training tasks in surgical simulation [1]. VR offers the possibility of practicing and planning procedures without risking and endangering the lives of the patients. The rigid endoscopy is a diagnosis technique in which a camera is inserted through a natural orifice of the human body thus, allowing the surgeon to see and navigate inside the cavities for planning and performing surgical procedures. For simulation purposes, the computer model of a patient is defined from medical imagery information and implemented to interact with an endoscope in a virtual environment. For increasing the realism in the simulation, it is necessary to integrate a haptic system when handling the endoscope, allowing the user to feel collisions during the navigation. The virtual collision sends a signal through its actuator, so the user senses it through its muscle-skeletal system, and the experience is complemented with visual feedback to enhance immersion and realism.

Some common haptic devices used for surgical purposes found in the market are: The Phantom Omni from SensAble Tecnology [2], which is a general purpose haptic system with a serial kinematic structure. The ProMIS system from Haptics Inc [3] allows a complete simulation for minimally invasive surgery. The LapMentor system from Simbionix [4] is used for laparoscopic surgery training with haptic feedback and interactive visualization. The STEPS system [5], simulates the movements of the endoscope and its access to the tumor by navigating with a force feedback joystick, however the navigation model proposed do not take into account tissue deformations for having a real haptic response. Finally the Sinus Endoscopy system [6] shows a

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 30–36, 2011.

realistic rendering using the GPU with a low load of CPU at interactive frame rates, but without haptic response.

This paper presents the design and implementation of a mechatronic system for simulating a virtual rigid endoscopy and its application during endonasal endoscopy [7]. The paper is organized as follows; in section 2 each component of the mechatronics system (mechanical, electronic and control subsystems) is presented. In section 3 the proposed prototype and its integration to the medical imaging software Nukak 3D are described. Finally, in section 4 conclusions and future work is presented.

2 Mechatronic System

The endoscopy simulator is composed of the subsystems presented in Fig.1. The actuator subsystem detects position changes in the effector, that can be sent to the virtual environment allowing the navigation of the generated volume using medical images from a Computed Tomography dataset. When a collision is detected between the user controlled effector and the virtual medical image volume, a control action is sent to the actuators for preventing the endoscope to remain in a collision state, thus giving tactile feedback based on the material and volume properties assigned.

The mechatronic system was designed using the endonasal endoscopy model proposed by Perez et al [8]. This model has four Degrees of Freedom (DOF) for measuring the position and orientation of the real endoscope prototype (roll, pitch, yaw and axial displacement) and it also has three DOF to generate haptic feedback (pitch, yaw and axial displacement) in the user's hand. The DOF are referenced to a pivot located at the entrance of the nasal cavity, as shown in Fig. 2.

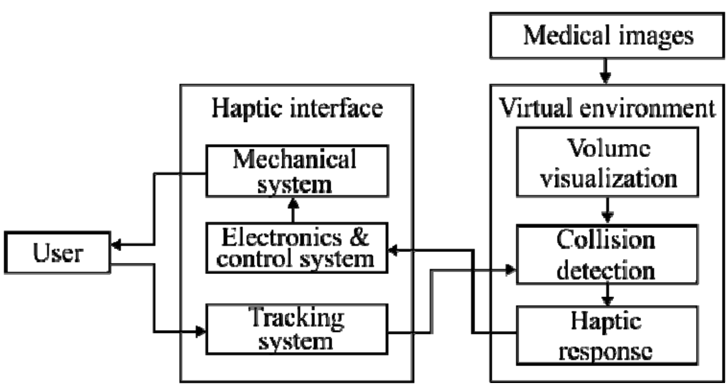


Fig. 1. Endoscopy simulator mechatronic system block diagram

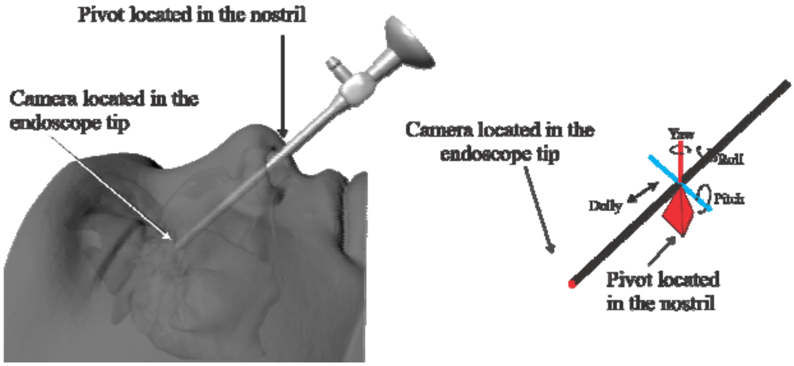


Fig. 2. DOF of mechatronic endoscope

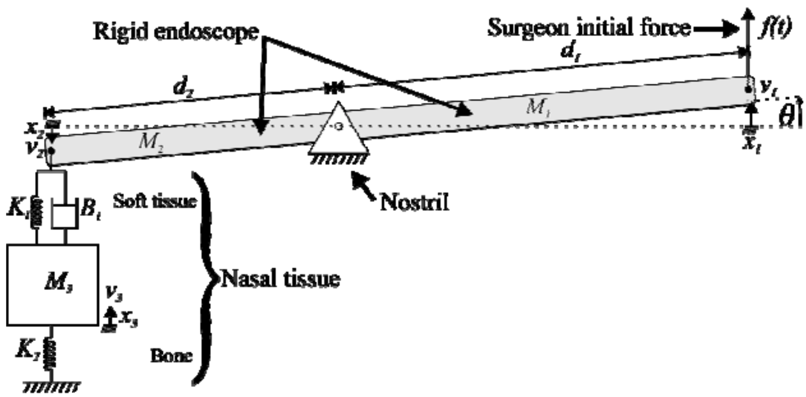


Fig. 3. Rigid endoscope model for virtual endoscopy

2.1 Rigid Endoscope Model

Rigid endoscope movements can be represented through as a 4 DOF mechanism mechanical joint which is analyzed by transforming the user's force applied over the nasal tissue. Fig. 3 shows the characterized endoscope model with a mass $M=M_1+M_2$, length $D=d_1+d_2$, and the nostril as a pivot point. Due to the anatomical restrictions of the nasal cavity, there is a limited space to rotate the endoscope therefore the θ angle is small and its behavior in terms of system's forces and velocities can be represented by (1) on the user side, (2) on the collision site and (3) between soft and osseous tissue.

$$0 = -f + m_1\dot{v}_1 + f_1 \ . \tag{1}$$

$$0 = -f_2 + m_2\dot{v}_2 + K_1\int v_2 - v_3 dt + B_1(v_2 - v_3) \ . \tag{2}$$

$$0 = K_1\int v_3 - v_2 dt + B_1(v_3 - v_2) + m_3\dot{v}_3 + K_2\int v_3 dt \ . \tag{3}$$

The elastic properties of the osseous tissue are modeled from the K_2 spring and M_3 mass and the elastic and damping properties of soft tissue are modeled from K_1 spring and B_1 damper, the viscous friction of the mucus.

2.2 Position and Orientation of the Tracking System

A mechanical structure for supporting the haptic system actuators (three Direct Current (DC) motors) and a real rigid endoscope prototype with magnetic position and orientation tracking (Flock of Birds (FOB) [9]) were designed. The receiver antenna is located at the free end of the endoscope, while the transmitter antenna is located on a flat surface within a 3ft radius of the receiver in order to accomplish the FOB specifications. The tracking system was configured to take 144samples/s at a communication rate of 115.2 kbps with the host computer, adequate for handling fast user´s movements. The designed mechatronics system is presented in Fig.4, where the mechanical structure, actuators, magnetic receiver and antenna can be seen.

2.3 Collision Detection and Force Feedback

The linear trajectory of the endoscope is affected by the rotations over the nostril, and the navigation limitations given the patient's anatomy. The camera range of motion is defined using the voxel information from CT images. The collision is detected when the destination point intensity is above a specified threshold obtained from the nasal cavity wall segmentation, stopping the camera motion.

Users can simulate the insertion of the endoscope into the nasal cavity and experience haptic feedback from the computed forces of the virtual endoscope collision against the walls of the nasal cavity using the mass-spring model with damping. These forces are obtained by solving the differential equations of the model and latter rendered onto the haptic device. The direction is found by calculating the difference between the position vector when the collision occurs and the previous position vector, this value is sent it to the electronic system that controls the DC motor motion information.

2.4 Electronic System

The electronic system governs the speed and rotation direction of each DC motor, it is responsible for transmitting a tactile feedback to the user through the endoscope. The system has four stages (see Fig. 5): the first one the electronic circuit with the computer, this is followed by a digital processing stage for translating commands received from the computer. The third one is the coupling stage, necessary for handling different voltages and currents in the same circuit, and finally a power stage to interact directly with DC motors.

The processing was implemented using a Microchip's PIC 18F4550 Microcontroller [10] with USB 2.0 communication protocol and signal generation obtained through a Pulse Width Modulation (PWM) for controlling motor velocity. The digital circuit and power stage coupling is implemented using an opto-isolator. These devices are capable of isolating voltage grounds from control and power stages. In the power stage an H-bridge circuit was used as the driver for controlling the rotation and the direction of the DC motors.

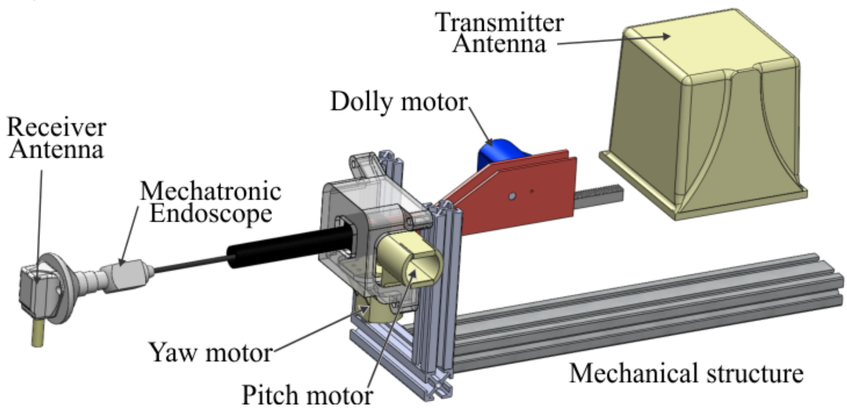


Fig. 4. CAD mechanical structure and components

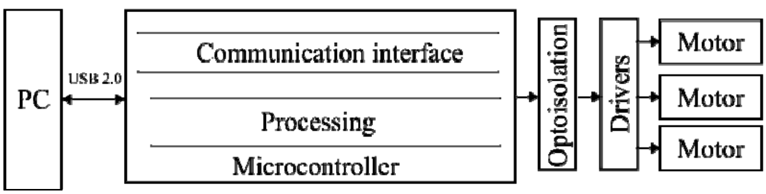


Fig. 5. Electronic system diagram

2.5 Computer Prototype

In order to visualize different anatomical structures, these must be segmented from the acquired dataset volume prior to the simulation. In some cases, the air-tissue interface can be hardly distinguished, due to the surrounding mucosa covering the thin bone walls, this is represented by a voxel intensity variation, and therefore thresholding methods are not good solutions to this problem. In the Nukak3D [11] software platform, this issue was solved through the levelset [12] method for segmenting the related anatomy. The images were rendered using texture mapping algorithm. Fig. 5 shows the mechatronic prototype and Nukak3D integration.

3 Test and Results

Several tests were performed in order to measure the accuracy of the prototype. The first test consisted in contrasting the tracking output values with physically measured positions, this comparison resulted in no differences between the data. The second test consisted in the calibration of virtual and real endoscope positions for obtaining proper camera movements within the virtual environment. For the electronic system, the test consisted in the verification of the outputs signals generated by the PWM and the current values supplied to the DC motors for its proper operation.

The tests were performed on a computer with Intel Xeon quad-core processor running at 2.6GHz, 4GB of RAM, Nvidia Quadro FX 1800 video card, and two datasets: 512x512x399 @ 16 bits and 512x512x100 @ 16 bits. The obtained frame rate of the render was between 15 and 20 frames per second (fps), enough to satisfy an interactive simulation without losing the sense of immersion. The haptic feedback was maintained in 1000Hz and it was not affected by the size of the medical dataset.

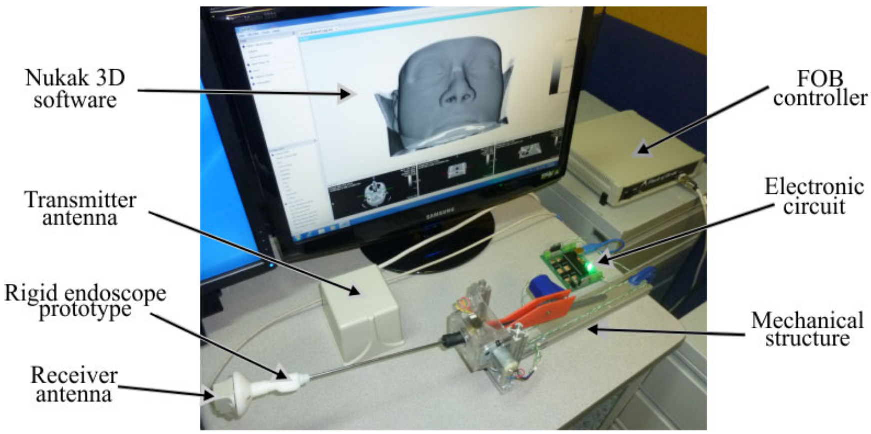


Fig. 6. VR and mechatronic rigid endoscope integration

4 Conclusions and Future Work

The successful integration of the mechanical, electronic and control components along with the software is the goal of any mechatronic project. In this paper mechatronic prototype for rigid endoscopy simulation based on endonasal endoscopy motions was presented. The haptics part of the prototype included the FOB tracking system for measuring the position and orientation of the mechatronic endoscope and the mechanical system with DC motors for force feedback. The dynamic model of the endoscope allowed the user to feel different tissue textures during a collision. The rendering speed between 15 and 20 fps depended on the size of the dataset, was considered sufficient to maintain a smooth simulation thus, avoiding seasickness during its use. We are currently running tests to determinate the impact of the prototype as a training tool in an academic environment.

Acknowledgement

This project was supported by the Research Division of Nueva Granada Military University ING379. Special thanks go to Alvaro Joffre Uribe for his good ideas and substantial support.

References

1. Li, S.: Research on application of virtual reality technology in teaching and training. In: Second International Conference on Computer and Electrical Engineering, ICCEE 2009, vol. 2, pp. 77–80 (2009)
2. Sensable Tech., Phantom omni haptic device, http://www.sensable.com/haptic-phantom-omni.htm
3. Promis, Surgical simulator, http://www.haptica.com/promis-surgical-simulator.html
4. Simbionix Ltd., Lap mentor, http://www.simbionix.com/LAP_Mentor.html
5. Neubauer, A., Wolfsberger, S., Forster, M.-T., Mroz, L., Wegenkittl, R., Buhler, K.: Advanced virtual endoscopic pituitary surgery. IEEE Transactions on Visualization and Computer Graphics 11(5), 497–507 (2005)
6. Kruger, A., Kubisch, C., Strauss, G., Preim, B.: Sinus endoscopy application of advanced gpu volume rendering for virtual endoscopy. IEEE Transactions on Visualization and Computer Graphics 14(6), 1491–1498 (2008)
7. Cappabianca, P., Cavallo, L.M., de Divitiis, E.: Endoscopic endonasal transsphenoidal surgery. Neurosurgery 55, 933–941 (2004)
8. Perez-Gutierrez, B., Martinez, D., Rojas, O.: Endoscopic endonasal haptic surgery simulator prototype: A rigid endoscope model. In: Virtual Reality Conference (VR), pp. 297–298. IEEE, Los Alamitos (2010)
9. Ascension Technologies, Flock of Birds, http://www.ascension-tech.com/
10. Microchip, PIC18F4550 Data sheet, http://www.microchip.com/wwwproducts/Devices.aspx?dDocName=en010300
11. Pinzón, A., Perez-Gutierrez, B., Romero, E.: Nukak3D, http://nukak3d.sourceforge.net/
12. Osher, S., Sethian, J.A.: Fronts propagating with curvature dependent speed: Algorithms based on hamilton-jacobi formulations. Journal of Computational Physics 79, 12–49 (1988)

Patterns of Gaming Preferences and Serious Game Effectiveness

Katelyn Procci, James Bohnsack, and Clint Bowers

Department of Psychology, University of Central Florida, 4000 Central Florida Boulevard, Orlando FL 32816
{kprocci,jbohnsack}@knights.ucf.edu, bowers@mail.ucf.edu

Abstract. According to the Technology Acceptance Model (TAM), important predictors of system use include application-specific self-efficacy, ease of use, and perceived usefulness. Current work with the TAM includes extending the assessment framework to domains such as serious games as well as how other typically under-researched factors, such as gender, affect technology use. The current work reports on how there are gender differences in both game playing behaviors as well as general game genre preferences, offers implications for serious game designers regarding the development of effective learning interventions based on these differences, and finally suggests avenues for future research in this area.

Keywords: gender differences, serious games, technology acceptance model, user preferences.

1 Introduction

Reports of the success of serious games have echoed across a variety of domains – from improving military operations to generating a more efficient workforce. As an entertainment form, it is generally accepted that more males play video games than females. For example, a study conducted by Bonanno and Kommers [1] found that male junior college students were far more likely than females to play games and that they played for longer, averaging 6.71 hours per week versus only 2.49 hours per week. The researchers found that if an individual reported playing games for 30 hours a week, there was only a 2.1% chance that they were female. As for game genre preference, females were more likely than males not to provide any game preferences at all, suggesting that they did not play games enough to have a preference. As the science behind serious games continues to develop, it is important to ask what the role of gender differences in gaming habits and preferences plays in how serious games may have differing levels of effectiveness based on prior gaming experience. For example, will male players of serious games learn more because of their own preferences and experiences with the medium? Why would this matter? We will be exploring this issue through the lens of the Technology Acceptance Model (TAM).

1.1 Technology Acceptance Model and Gender

The TAM, developed by Davis and colleagues [2, 3], seeks to explain why users choose to interact with technology. Critical to this model are the perceptions of the

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 37–43, 2011.

users themselves. The use of a given piece of technology is driven by the intention to use, which in turn develops if the users feel that the technology is both useful and easy to interact with. This is tied to their own self-efficacy, which is the extent to which they believe that they are able to achieve specific goals.

The TAM has been expanded to include earlier predictors of intention to use in information systems. A study by Yi and Hwang [4] confirmed the importance of enjoyment, learning goal orientation, and application-specific self-efficacy when determining the likelihood of students' use of Blackboard, a commonly used internet-based educational content delivery system. Interestingly, they found that application-specific self-efficacy was one of the strongest predictors of use, even beyond that of intention to use. Thus, how capable the users felt of being able to use the specific technology was extremely important.

As such, the authors point out the necessity of examining the ways in which individual differences, such as gender, may ultimately affect technology acceptance and use. They cite the work of Gefen and Straub [5], which was the first to highlight how gender is often ignored when applying the TAM to instructional systems. The researchers found that when comparing the attitudes of email usage between the genders, there were differences in social presence, perceived usefulness, and application-specific self-efficacy. However, they failed to find this reflected in actual e-mail use. Regardless, many researchers continue to urge further exploration into the role of individual differences such as gender in all technology acceptance models [6]. Perhaps when given a specific task that is traditionally oriented toward one of the genders, such as with video games, an individual’s application-specific self efficacy and perceived usefulness could potentially differ based on whether they are male or female, ultimately resulting in different patterns of use.

1.2 Applying the TAM to Serious Games

The TAM has also been applied to serious games. In his unpublished dissertation, Evans [7] was able to successfully apply a modified version of the TAM to the use of a serious game, in which he found that earlier predictors such as learning goal orientation, enjoyment, ease of use, application-specific self-efficacy, and perceived usefulness predicted behavioral intention which then predicted both actual use as well as improved learning outcomes. As the goals of a serious game are to achieve knowledge retention, this is a very important finding. Since research on the TAM suggests that simply feeling capable of playing a specific serious game will be an important predictor of use and also given the extant differences between genders in the gaming domain, it seems pertinent to examine the ways in which gender can affect the predictors of use and overall effectiveness of serious games. Specifically, we feel that examining gender-based differences in preferences for and playing behaviors of entertainment games may provide insight into the ways that serious games may be perceived differently by males and females as well as how that might affect application-specific self-efficacy and, eventually, alter learning outcomes. This will reveal implications for the way serious games are designed.

2 Gender Differences in Game Playing Preferences and Behaviors

Speaking broadly, males seem to prefer action-oriented games and first-person shooters while females are more likely to play and enjoy simpler, more puzzle-like casual games. To explain this, the literature relating to the gender differences in gaming preferences is split between two distinct camps – the neurocognitive and the psychosocial.

2.1 Neurocognitive Basis for Preferences and Behaviors

There are some very specific, generally accepted neurocognitive gender differences. Males tend to have an advantage over females in 3D mental rotation, spatial cognition, and visuospatial reasoning. Females have the advantage when tasks involve the need for perceptual speed and fine motor skills and prefer concrete, repetitive activities [1]. Most of the popular commercial video games rely heavily on mental rotation abilities in order to play them. For example, shooters are often male favorites and make use of 3D mental rotation, spatial perception, and targeting, all of which males excel in, while the preferred games of females tend to be casual puzzle games, which require color memory and verbal skills, both of which females are traditionally more adept [8]. Bonanno and Kommers [1] suggest that these natural tendencies translate into game preferences in that individuals will self-select games to play that make use of the skills in which they are naturally better.

2.2 Psychosocial Basis for Preferences and Behaviors

Other researchers suggest that the gender differences are based in psychosocial differences. Hartmann and Klimmt [9] found that females, in general, seemed to dislike media and games with negative or sexually-suggestive stereotypes, with violence, and those that placed them into a competitive role. They also found that most commercially available games contained females in sexually suggestive roles, violence, and competition, which may be the reason behind why males are so much more interested in gaming then females as they are not as bothered by these elements. Glazer [10] agrees with the assertion that females are interested in games but the majority of games on the market do not cater to their specific tastes. His research found that games females enjoy feature open-ended exploration, involve team-like play as opposed to actual competition, and promote networking and personal interaction, which are very unlike the first-person shooters and other physical games that seem to dominate the gaming market.

2.3 Present Study: Gender and Game Preferences

In summary, the literature suggests that males prefer action-oriented and competitive games while females are more likely to enjoy playing puzzle-based, casual, and cooperative games. While these preferences can be explained from both the neurocognitive and psychosocial perspectives, it is evident that they exist and have consequences in the real-world as reflected in actual gameplaying behavior.

We have found a similar pattern of preferences and playing behaviors in our own studies of games. We collected survey responses from 760 undergraduate students from a large southeastern university concerning their game playing behaviors and preferences. From these, 245 were males and 515 were female. Respondents reported a mean age of 19.88 years (*SD* = 3.83). When analyzing the data, we uncovered several interesting gender differences. Of those who responded, 71.43% of males qualified as gamers, which we defined as those who played video games at least twice a week and/or for an average of at least five hours per week, while only 27.57% of females qualified. Males played an average of 9.01 hours of games per week (*SD* = 10.95) while females played an only an average of 2.84 hours (*SD* = 3.92). These findings are consistent with the previous research. We asked our respondents to rate on a 5-point Likert scale their preferences for a variety of genres as well as list the title of their favorite game, which were coded by genre. We conducted a series of t-tests at the .05 level for each genre to determine whether differences in preferences by gender were statistically significant. The differences were clear (see Table 1).

Table 1. Mean genre ratings and reported favorite games by gender

Genre	Males	Females	*t*	Fav. (M)	Fav. (F)
Strategy	**3.45 (1.19)**	2.52 (1.23)	9.745*	3.72%	0.08%
Fantasy/RPG	**3.18 (1.37)**	2.64 (1.35)	5.036*	19.07%	9.26%
Action-Adventure	**4.00 (1.08)**	2.77 (1.08)	12.115*	6.51%	8.17%
Shooter	**4.27 (1.08)**	2.79 (1.37)	14.550*	45.12%	7.63%
Fighter	**3.45 (1.19)**	2.73 (1.37)	6.886*	3.26%	2.72%
Simulation	3.04 (1.16)	**3.58 (1.34)**	-5.246*	0.47%	12.81%
Classic Arcade	3.66 (1.01)	**4.34 (0.95)**	-8.814*	0.47%	9.81%
Card Games	2.79 (1.22)	**3.42 (1.29)**	-6.229*	1.40%	0.27%
Quiz Games	2.66 (1.22)	**3.82 (1.14)**	-12.472*	0.00%	0.82%
Kids Games	1.80 (0.96)	**2.49 (1.16)**	-7.924*	0.00%	0.54%
Sports	**3.47 (1.47)**	2.58 (1.31)	8.176*	17.67%	4.63%
Racing	3.68 (1.12)	3.55 (1.23)	1.347	0.47%	4.90%
Puzzle	3.09 (1.43)	**4.14 (1.05)**	-12.491*	1.40%	22.89%
Massively Multiplayer Online	**3.03 (1.43)**	2.09 (1.20)	9.184*	0.00%	6.82%
Competitive Multiplayer	**4.08 (1.24)**	2.36 (1.26)	17.383*	0.00%	1.36%

Notes: $*p < .001$; standard deviations listed in parentheses. For statistically significant differences, the higher scores, which indicate more enjoyment, are in bold typeface; Fav. refers to percentage of males then females who endorsed that particular genre as their favorite game of those that were able to provide a favorite game (males $n = 215$, females $n = 367$). Percentages are rounded to two decimal places, resulting in slightly less than 100%.

The highest rated game genres by males were the action-adventure, shooter, and competitive multiplayer games, while females tended to prefer the classic arcade, quiz, puzzle, and card games. Regarding actual favorite games, by and large, females were far more likely than males to not list any games that they enjoyed playing (28.73% females as compared to 12.24% of males), similar to previous findings [1]. Of those that did list a favorite game, the preferred genres as indicated by favorite game for males are by far the more popular and traditional games such as shooters, fantasy and role-playing games, and sports games (81.86% of all favorite genres) as compared to females who endorsed more casual games, such as puzzle games, simulation games, and classic arcade games (45.51% of all favorite genres).

We also asked respondents to indicate whether or not they played and enjoyed playing competitive and cooperative multiplayer games. Males were far more likely to play multiplayer games than females, with 85.31% of males compared to only 50.39% of females having reported playing them. Also, 45.71% of males preferred multiplayer to single-player play while only 19.18% of females indicated the same. If given the choice between competitive and casual/cooperative multiplayer games, males were divided evenly between preferring either gameplay type (53.50% preferred competitive) while females were only slightly more biased toward cooperative multiplayer gaming (58.48%). Males were far more likely to play multiplayer games online (79.84% of males as compared to 26.26% of females), but also tended to endorse playing with friends either online or in the same location together (53.01% vs. 46.91%) while females overall preferred to play with friends in the same room (76.65%).

3 Implications and Future Directions for Serious Games Designers and Researchers

Overall, the data collected from our sample was in support of the current body of literature. Armed with this knowledge, the value of these numbers can be seen as we attempt to extrapolate meaning. Specifically, as more serious games are being developed to be used as educational tools for those in high school and early in students' undergraduate careers, a snapshot of preferences and behaviors may prove to be useful. The inherent mechanism driving the push to use serious games is that they are just that – games. If individuals enjoy a learning intervention, they will engage with it more, and thus hopefully the repeated exposure will result in improved retention. Therefore, understanding and applying player preferences is vital, especially if believing whether one can successfully play a serious game based on personal experience is a factor driving use and eventual effectiveness.

These patterns, regardless of underlying cause, can result in developing stereotype threat. Stemming from the work of Steele and Aronson [11], stereotype threat has been defined as when a negative stereotype about a certain group exists and actually affects the performance of a member of that group if they are aware of the stereotype. Stereotype threat based on gender is no exception [12, 13]. Here, could females being aware that they are not the typical gamer actually result in decreasing the amount of information that can be learned from a serious game? This also may be detrimental to the application-specific self-efficacy of female serious game players. According to the TAM, it follows that a player's feelings as to whether they can successfully play a specific type of game as well as whether or not it is easily accessible may in fact affect the actual learning outcomes gained as a function of play.

3.1 Designing Serious Games for Both Genders

With gender differences that are so clearly present, it becomes important to consider the role a game's genre will play in the overall effectiveness of a given serious game especially if the intended audience is composed of both males and females with differing levels of game experience. For example, if a designer creates a competitive multiplayer real-time strategy serious game to educate students about business resource management, the male students may feel more secure and comfortable with playing the game because they might be more familiar with the expected gameplay mechanics and controls. This will result in a greater sense of application-specific self-efficacy, resulting in increased use and improved learning outcomes over those who are not as familiar with this particular style of game. Also, from the perspective of cognitive load, if mental resources are expended on attempting to understand and use a learning tool, only a limited amount of cognitive reserve can be used to actually learn the target content. Those that are more familiar with a specific genre will already know what to expect of the game and excel, while those who are not must first expend the extra mental effort to master the games' mechanics before being able to focus on the educational content. If a given genre is more well-known by only one of the genders, this potentially will result in playing differences and ability that may affect learning outcomes. Thus, something simple yet important such as choosing the appropriate genre for a serious game is a design decision with a very broad impact.

Despite the important differences, it is not feasible to develop two versions of every serious game to cater to gender differences. Given the limited resources available to serious game designers, the focus should instead be on designing a gender-inclusive game. By examining both the literature as well as our data, we can provide some tentative suggestions for future game development efforts that are also avenues for future research.

By mixing elements that cater to both genders' preferences, it may be possible to capture the interest of both. One way to create games that are gender-inclusive could be to create a social environment that is collaboratively competitive. As for cooperative play, our findings suggest that while males have no preference, females prefer to play in the same room with friends. For serious games, this suggests that designers can create games that can be played on the same console or computer together for the potential benefit of both genders. Not only this, from a return on investment perspective, this means fewer pieces of equipment will be necessary to play such games in the educational setting.

Some research suggests that being able to customize the physical appearance of in-game player characters improves females' interest in Massive Multiplayer Online games (MMOs) [14,15]. While we did not find MMOs to be highly rated by females, perhaps including optimal character customization may increase female interest in the serious game without alienating the male players. Also, if you do choose to create a game that is less well-liked or relatively unfamiliar to females, such as a strategy game, you may want to consider taking extra time to develop in-depth gameplay tutorials as well as provide positive and supportive feedback during play. You can also lessen the steepness of the difficulty curve early-on in gameplay to improve the players' sense of application-specific self-efficacy.

Future work should also explore the ways in which these preferences directly affect learning outcomes in serious games. Pertinent research questions could include whether experience with a given genre results in improved in-game performance as well as increased retention, if gamers in general are able to learn more from serious games, as well as the ways in which gender and gaming behaviors interact with genre and gameplay mechanisms to affect learning outcomes. This work, along with more research into related questions concerning individual differences, is needed to generate a framework for game development that will improve the overall effectiveness of games for both males and females while maintaining a reasonable return on investment.

References

1. Bonanno, P., Kommers, P.A.M.: Gender Differences and Styles in the Use of Digital Games. Educ. Psychol. 25, 13–41 (2005)
2. Davis, F.D., Bagozzi, P., Warshaw, P.R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Manage. Sci. 35, 982–1003 (1989)
3. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quart. 13, 319–340 (1989)
4. Yi, Y.M., Hwang, Y.: Predicting the Use of Web-Based Information Systems: Self-Efficacy, Enjoyment, Learning Goal Orientation, and the Technology Acceptance Model. Int. J. Hum.-Comput. Int. 59, 431–449 (2003)
5. Gefen, D., Straub, D.W.: Gender Differences in the Perception and Use of E-mail: An Extension to the Technology Acceptance Model. MIS Quart. 21, 389–400 (1997)
6. Sun, H., Zhang, P.: The Role of Moderating Factors in User Technology Acceptance. Int. J. Hum-Comput. St. 64, 53–78 (2006)
7. Evans, A.W.: Learning for the Next Generation: Predicting the Usage of Synthetic Learning. Doctoral dissertation, University of Central Florida (2009)
8. Sherry, J.L.: Flow and Media Enjoyment. Commun. Theor. 14, 328–347 (2004)
9. Hartmann, T., Klimmt, C.: Gender and Computer Games: Exploring Females' Dislikes. J. Comput-Media Commun. 11, 910–931 (2006)
10. Glazer, S.: Video Games. CQ Res. 16, 937–960 (2006)
11. Steele, C.M., Aronson, J.: Stereotype Threat and the Intellectual Test-Performance of African-Americans. J. Pers. Soc. Psychol. 69, 797–811 (1995)
12. Spencer, S.J., Steele, C.M., Quinn, D.M.: Stereotype Threat and Women's Math Performance. J. Exp. Soc. Psychol. 35, 4–28 (1999)
13. Aronson, J., Lustina, M.J., Good, C., Keough, K.: White Men Can't Do Math: Necessary and Sufficient Factors in Stereotype Threat. J. Exp. Soc. Psychol. 35, 29–46 (1999)
14. Dede, C., Ketelhut, D., Nelson, B.: Design-Based Research on Gender, Class, Race, and Ethnicity in a Multi-User Virtual Environment. Paper presentation. American Educational Research Association Annual Meeting (2004)
15. Turkay, S., Adinolf, S.: Free to be Me: A Survey Study on Customization with World of Warcraft and City of Heroes/Villains Players. Procedia - Soc. Behav. Sci. 2, 1840–1845 (2010)

Serious Games for the Therapy of the Posttraumatic Stress Disorder of Children and Adolescents

Rafael Radkowski[1], Wilfried Huck[2], Gitta Domik[3], and Martin Holtmann[2]

[1] Heinz Nixdorf Institute, Fürstenallee 11, 33102 Paderborn, Germany
[2] LWL-University Clinic Hamm, Heithofer Allee 64, 59071 Hamm, Germany
[3] University of Paderborn, Fürstenallee 11, 33102 Paderborn, Germany
rafael.radkowski@hni.uni-paderborn.de, domik@uni-paderborn.de, {wilfried.huck,martin.holtmann}@wkp-lwl.org

Abstract. The posttraumatic stress disorder (PTSD) is a mental-health problem that can emerge after a delayed reaction of a person to a traumatic incident. A common therapy is the so-called exposure therapy. However, children and adolescent cannot be treated with a common therapy. In this paper we describe a serious game for the therapy of a PTSD by children and adolescent. Objective of this paper is to introduce a concept for the game development and a method to balance the game. It is based on a so-called impulse hierarchy. The game prototype respectively the utilized concept and methods have been tested with healthy test persons. The results are a strong indication for the effective of the developed game concept and the usefulness of the key principles.

1 Introduction

The posttraumatic stress disorder (PTSD) is a mental health problem that can emerge after a delayed reaction of a person to a traumatic incident. This incident leads to an exceeding threat of the relevant person. It arouses a sense of deep desperation, strong fear, and helplessness. A traffic accident is such an incident. Involved people are not able to manage their life anymore on their own. They avoid the traumatic situation, retire themselves from their social environment, suffer sleeplessness, great nervousness, and other mental disorders. A behavior therapy can help the involved people. A common therapy is the so-called exposure therapy. This therapy confronts the patient with his / her individual trauma. It includes a repeated relive of the traumatic situation. Normally, the repeated relive is done narrative or in-vivo; at the place where the trauma occurs. This therapy enforces an emotional rehabilitation and leads to a systematic desensitization.

Beside the classical therapy, virtual reality exposure therapies (VRET), game-like applications and serious games are explored for several years. These therapies displace the patient into a virtual world and confront the patient with virtual stimuli. Games add a story that enhances the reflection of complex mental health problems. Altogether, the research in this field demonstrates the effectiveness of this therapy method.

However, children and adolescents cannot be treated in the same way. They have a different comprehension of the PTSD and a lack of insight into the difficulties that the

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 44–53, 2011.

mental disorder cause. In all, they need to be motivated for the therapy. However, serious games are in the focus of research for this target group in order to treat mental health problems [1], [2]. In our definition, serious games are computer games that impart knowledge, train capabilities as well as enable them to apply this knowledge meaningful. The knowledge is transferable into the "real world", the player becomes more capable of acting. During the therapy the children and adolescents play the serious game. Thus, they become part of a metaphorical adventure. Indirectly, they deal with different important questions regarding their mental health problem.

In this paper we describe key principles for the design of a serious game for the therapy of PTSD. Furthermore, we introduce a method to keep the game in balance between fun and fear. Both methods facilitate third persons to develop a serious game for the desired purpose. The paper is structured as following. First, the related work is described. Then the key principles and our game concept are introduced, followed by the explanation of the developed method. Afterwards, the tests are presented. The paper closes with a conclusion and an outlook.

2 Related Work

Serious games, commercial games, and game-like applications have found their way into the focus of therapists. Commercial computer games were the first games therapists used in therapy (e.g. [3], [4], [5]). The games have been used to treat different mental health problems. A patient, child or adolescent, played a game during a therapy session. The therapist observes his/her patient while he/she is playing and notes the patient his/hers behavior. The game serves as a context for discussion between therapist and patient. It provides a structure for a therapy session and supports the building of an effective patient-therapist working relationship [6].

A major concern regarding commercial computer games is their content, which limited its use in psychotherapy [7]. Commercial computer games, their graphical interface, the sound, and the game play are designed to evoke specific emotions such as fear, excitement, and humor [8]. These emotions interfere with the emotions a therapist needs to evoke for treatment [1].

To face this concern serious games for the treatment of mental health problems get into academic focus. An advantage of serious games is their motivating effect to children and adolescents. It makes fun to play a game. Serious games exploit this fun factor for therapy purpose: The treatment of PTBS and other mental health problems based on learning theories. And fun advances the learning process.

Different serious games have been designed for different mental health problems and for different therapy methods. But the main principle of all games is similar: The player, normally the patient, assumes the role of a virtual character and needs to handle different quests. The patient experiences a metaphorical story in the game. Thereby s/he deals with different questions and problems regarding the certain mental health problem. In detail the mechanisms differ that cause a therapeutic effect.

One of the first serious games in this field is the game Personal Investigator (PI) [9], [10]. PI has been developed for the treatment of emotional disorders. The patient assumes the role of a detective. Objective of the game is to become a master detective. During the game the player encounters different challenges and problems.

The solution of these problems helps the patient to reflect his/her problems. The game is in a clinical trial.

A second game is Treasure Hunt [11]. It is a game for the behavior therapy of children and adolescents with anxieties or depressions. The patient assumes the role of a child and needs to help the captain of a treasure ship reading a treasury map. For this purpose the patient has to solve different tasks that allow a reflection of the mental health problem.

Virtual Classroom is a game for children with an attention deficit disorder [12]. The game presents a virtual classroom including tables, chairs, a blackboard, virtual classmates, and a virtual teacher. The patient himself/herself is a scholar. During the game nuisance-causing effects like sound are brought in. The reactions of the patient are recognized and reflected.

There are still a few further games, which become effective utilizing the same major principle. Earthquake in Zipland is a serious game that should help children to elaborate the divorce of their parents [13]. A patient assumes the role of a moose and has to solve different quests. In 2010 the German game "Zappeliz zaubert" (Zappeliz perform magic) was released [14]. The game has been designed for the treatment of an attention deficit disorder.

In summary, several results have been presented:

- Serious games facilitate the treatment of mental health problems. The shown studies are a strong indication for their effectiveness.
- Serious games help the therapist to develop a working relationship to a patient. Many children have a problem to talk to a therapist due to several prejudices.
- Finally, serious games motivate children and adolescents. There are strong indications that this effect supports the therapy: Therapy is learning and motivation improves learning [15].

However, the serious games have not been developed for the treatment of a post-traumatic stress disorder (PTSD). Applications for the treatment of PTSD have been developed for adults ([16], [17], [18], [19]). Unfortunately, these applications are not designed as games. Thus, the results cannot be transferred without any further evaluation due to the fact that children and adolescents have a different understanding of the therapy. Finally, they need games to be motivated for therapy.

3 Game Concept and Principles

A serious game for the treatment of a PTSD needs to follow different principles than serious game for other mental health problems. The difference is: to treat a PTSD the patient must be confronted with his/her trauma, with a trigger stimulus that causes anxiety and unease. On the other side the serious game is used to evoke fun and gratification; both facilitate the treatment. If to much anxiety arise, the game causes the opposite effect. To face this challenge a set of key principles and an effective hierarchy are introduced. Both allow keeping balance between fear and fun.

3.1 Game Design and Key Principles

During the collaborative work on the serious game, a set of key principles occurs that guided the creation of a virtual game environment. We define a key principle (KP) as a design rule that directs the structure and layout of the virtual environment as well as the usage of trigger stimuli and their virtual presentation. The key principles have been founded after the observation and discussion of many therapy sessions and the development of a game prototype. **Fig. 1** shows a screenshot of the players view on the virtual game environment of this prototype.

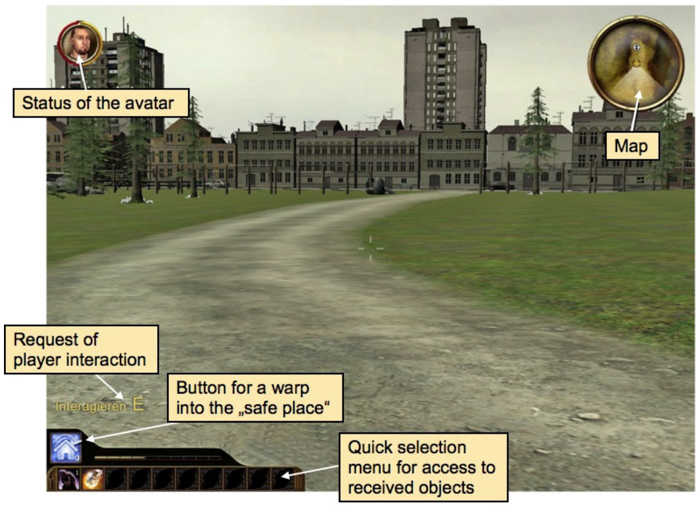


Fig. 1. The players view on the virtual game environment

The patient controls a virtual avatar that s/he plays from a first person view. The entire game is created like a normal computer game. On the main screen the player founds a status display of the avatar, a small map indicating the position of the avatar. A quick selection menu gives access to different objects, which the player receives as gratification during the game. The game is controlled by mouse and keyboard.

The basic game design needs to correlate with the individual experienced trauma. Everybody who suffers on a PTSD has experienced an individual event like a car accident. Thus, the trigger stimulus, the entire situation is individual and has to be designed individually. For instance, if a child has been involved in an accident at a crossroad inside a town and the car was red (this can be a trigger), the virtual world also needs to show a virtual crossroad and a red car. This is our major key principle for an effective serious game. Of course, the challenge is to identify the trigger stimuli. Mostly, also the children and adolescents do not know all trigger stimuli at the beginning of a treatment. To face this problem the game design, respectively the design of the virtual world needs to be adapted during the therapy.

In addition to this major principle, the further key principles define guidelines for the entire structure and setup of the virtual environment.

KP1 - The virtual environment needs to be separated into a mediation zone and a confrontation zone. Fig. 2 (left) shows an overview of the current virtual world realized in our prototype. On the top of the figure a red border indicates a so-called confrontation zone. In this zone only, the patient is confronted with his/her trigger stimuli. Beyond this zone the green border indicates a mediation zone. In this zone, the patient can talk to other non-player characters, etc. The separation into zones needs to communicate to the patient. The zones allow the patient to control the therapy, one important factor for its effectiveness. If the patient is ready to face a stimulus, s/he can walk into the confrontation zone. By all means, this requires an action of the patient (walk) that allows control.

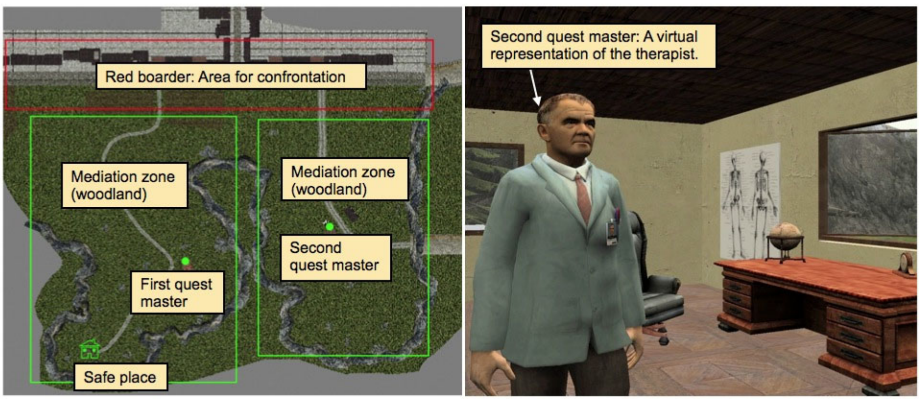


Fig. 2. An overview of the current virtual world (left), a quest master offers the player a set of quests: tasks that force the player to deal with his / her problem (right)

KP2 - Free choice of quests. The patient determines the "speed" of the therapy and not the therapist. Thus, s/he also has to choice between several quests. A quest is a task inside the game that the player has to perform. Most of them are search & collection quests as well as protecting quests. In each quest, the player deals with different aspects regarding the PTSD. For instance: "Why is it impossible to avoid the emotion of fear?", or "What trigger stimuli are to avoid?". However, the patient has to get the feeling of choosing the quest. S/he has to decide when s/he is ready to deal with a certain topic.

KP 3 – Gradual increase of stress. At the beginning of the game, in the first therapy session, the player should get the choice between "easy" quests. Each quest results in an increase of stress. The first task should result in a marginal increase of stress. With progressing therapy, the quests have to become more challenging. A practical solution to keep this under control is to establish a central point inside the virtual world and to assign quests with respect to the distance to that place. **Fig. 2**, left, shows this principle: The player starts at the so-called "safe place". The first quest master assigns only simple quests to the player. The second quest master assigns the challenging quests. Finally, the therapist should activate the last stimuli manually. In the case of a car accident, the patient has to see a virtual accident (Fig. 3, right).

KP4 - Individualization facilitates a central point. The central point inside the virtual world should be the so-called "safe place". The "safe place" is a psychological metaphor for a state of safety. The common therapy utilizes pictures, photos, or thoughts as safe place. In the game it has to be a 3D model at least. Fig. 3, left, shows our interpretation of this place. It is a building inside the game environment. To establish this place as a central point of the game, as a safe place, the player needs a strong relation to this place. This is possible by individualization. The player needs the possibility to design this place as s/he wish. Furniture, pictures, the design of the wall, and everything else has to meet the expectations of the player. This results in an emotional relation and causes a feeling of safeness.

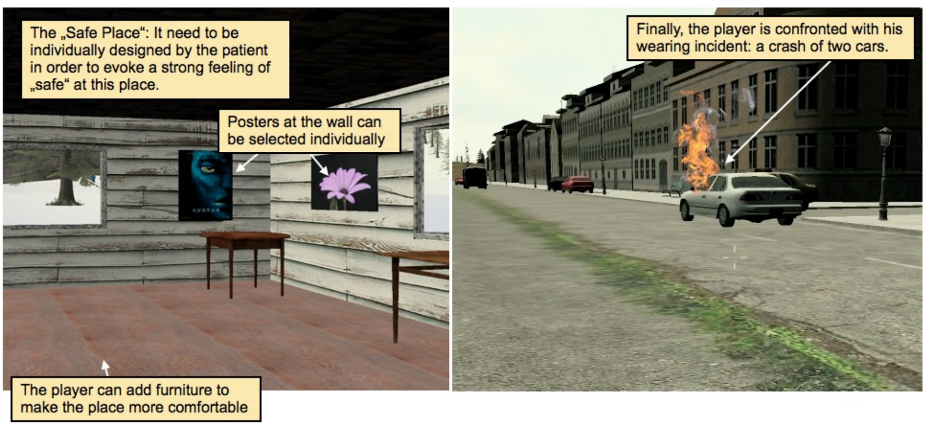


Fig. 3. The "safe place" (left), and a crash of two cars as final trigger stimulus (right)

KP3 - Gratifications have to be objects the patient can use immediately. After passing a quest the player must receive a gratification immediately. The children and adolescents need this gratification to evoke a feeling of satisfaction. This enhances the therapy. However, to realize this effect an object has to be deliver that can be used immediately in the game. An armor, a protective shield, and other things. Its effect is mitigates when game points or collectibles are rewarded.

3.2 Effective Hierarchy

To keep balance between anxiety and fun in the game an effective hierarchy has been developed. The effective hierarchy allows to keep track of the different stimuli in the virtual world and to balance the challenges in the game, caused by the quests. Figure 4 shows a section of the suggested effective hierarchy and its utilization.

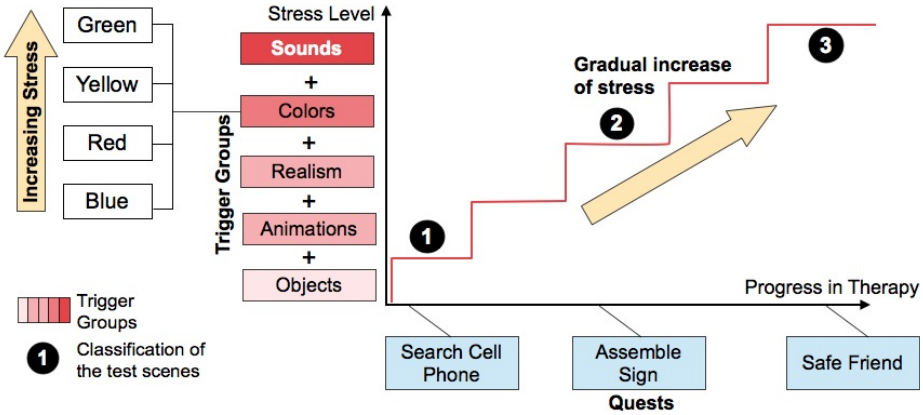


Fig. 4. Schematic overview of the effective hierarchy and its utilization

The basic of this effective hierarchy are a group of trigger elements, which are sorted by their ability to cause stress. The series is: objects (streets, houses, persons, trees, etc.), animations, realism, colors, and finally sounds. Basic elements are objects. To get a relaxing world, it has to be created using static virtual objects only. To increase the stress level animations have to be added. Then the realism of the scene should be increased by including details. The last two trigger groups are colors and sounds. If a scene contains all triggers, it becomes a stressful scene.

Each trigger group includes a set of graphical attributes that realize this trigger group. For instance, the trigger group color includes the colors blue, red, yellow, and green. The triggers are ordered by its ability to cause stress.

The order of the graphical attributes and sounds depends on the experiences of every patient. Thus, the therapist has to adapt this series for each patient. The therapist starts with a predefined template. During his/her talks with a patient s/he has to determine trigger stimuli (The therapist has to do this during a common therapy, too) and to reorder the stimuli / graphical attributes of the trigger groups.

In order to utilize this hierarchy it is linked to the different quests. In figure 4 the quests are indicated on the abscissa. From the left to the right the quests become more challenging. After a quest is designed, the included trigger stimuli (objects, animations, realism, color, and sound) need to be analyzed and their level of stress has to be rated. By this rating they can be assigned on the abscissa.

In the game a gradual increase of stress has to be realized. By the use of the effective hierarchy, we have an indication for the mental challenges a quest causes and are able to offer the patient a set of adequate quests.

4 User Tests

The key principles and the effective hierarchy need to be tested to get an indication for their effectiveness. Therefore, a test with healthy children (8 children) has been carried out. To compare this group, a group of adults (6 adults) has also been tested.

4.1 Test Setup

Fig. 5 (left) shows an image of the test setup. A test person observes three different, pre-defined scenes of the game on a large powerwall. The selection of test scenes is indicated in figure 4 by the numbers in the chart. The tests should increase the stress level of the test persons. Thus, scenes on a lower, an average, and a high-assumed stress level have been defined. We have assumed that the stress level of the test persons should increase, while observing the scenes.

To measure the stress of a test person and his/her "fear" the State-Trait Anxiety Inventory (STAI) has been used [20]. STAI provides a questionnaire with questions for 20 different items. The analysis of this questionnaire allows determining a stress level as quantitative value. The stress level has been determinate after each scene. A test person has observed in scene. After each scene the person has been asked for an appraisal of his/her current state.

In addition to the questionnaire, the pulse rate of each person has been measured. For this purpose a Polar pulse watch have been utilized.

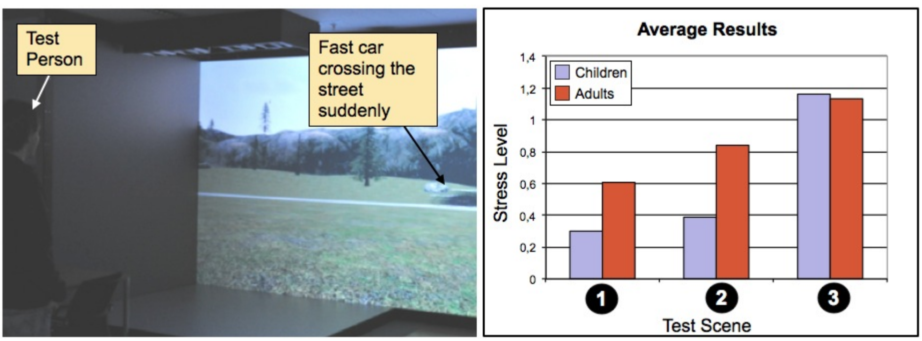


Fig. 5. An image of the test setup (left), the received results (right)

4.2 Results and Discussion

Fig. 5 (right) shows the results. The ordinate shows the stress level. The abscissa presents the three tests and the results. The bars show the average stress level of the three tests, calculated due to the STAI proceeding. The results are separated. The results of the children are colored mauve, the results of the adults are colored orange. The numbers (1 to 3) indicate the tests.

As expected the stress level increases due to the predicted stress level. The addition of trigger stimuli, according to the effective hierarchy, causes the expected behavior: From scene one to scene three the stress level increases. There are differences between the stress level of the adults and the stress level of children. Finally, both incline to the higher stress level. The lower stress level of the children at the scene one may base upon their familiarization with computer games. The pulse measurement supports these results in general. However, the gradient of the pulse rate is not as high as the gradient shown in figure 5.

5 Conclusion and Outlook

The contribution introduces the use of serious game for the treatment of a PTSD by children and adolescents. To design a virtual environment for this type of game, we introduce a set of key principles. In addition, to keep the challenges of the game in balance, an effective hierarchy was introduced. This effective hierarchy should make the scenes and their effect predictable.

To determine their usefulness a user test has been carried out. This test and the developed prototype enable three statements:

First, the key principles are practicable for game design. They provide a guideline for the arrangement of composition of 3D models like streets, buildings, woodlands, and quests, respectively the quest objects. The key principles are not fix but they are a good starting point.

Second, the effective hierarchy and the ordered trigger elements are a useable method to design a concrete therapy. However, they are not finally evaluated. But the results of the test are a strong indication that they facilitate the generation of virtual scenes with a controllable and predictable stress level.

Third, it is possible to design a serious game that keeps balance between anxiety and fun. Thus, in future the game will facilitate the therapy of PTSD.

Our next step is to extend the game. More levels and a larger map are necessary in order to keep a child busy for a several hours. Furthermore, we prepare a controlled study with children that show pre-clinical symptoms of a PTSD.

References

1. Ceranoglu, T.A.: Video Games in Psychotherapy. Review of General Psychology 14(2), 141–146 (2010)
2. Kato, P.M.: Video Games in Health Care: Closing the Gap. Review of General Psychology 14(2), 141–146 (2010)
3. Kappes, B.M., Thompson, D.L.: Biofeedback vs. video games: Effects on impulsivity, locus of control and self-concept with incarcerated juveniles. Journal of Clinical Psychology 41(5), 698–706 (1985)
4. Gardner, J.E.: Can the Mario Bros. help? Nintendo games as an adjunct in psychotherapy with children. Psychotherapy: Theory, Research, Practice, Training 28, 667–670 (1991)
5. Kokish, R.: Experiences using a PC in play therapy with children. Computers in Human Services 11, 141–150 (1994)
6. Coyle, D., Doherty, G., Sharry, J.: An evaluation of a solution- focused computer game in adolescent interventions. Clinical Child Psychology and Psychiatry 14, 345–360 (2009)
7. Zelnick, L.: The computer as an object of play in child treatment. Journal of Infant, Child, and Adolescent Psychotherapy 4, 209–217 (2005)
8. Bertolini, R., Nissim, S.: Video games and children's imagination. Journal of Child Psychotherapy 28, 305–325 (2002)
9. Coyle, D., Matthews, M.: Personal Investigator: a Therapeutic 3D Game for Teenagers. In: Computer Human Interaction 2004 (CHI 2004) - Social Learning Through Gaming Workshop, Vienna, April 25-29 (2004)

10. Coyle, D., Doherty, G.: Clinical Evaluations and Collaborative Design: developing new technologies for mental healthcare interventions. In: Proceedings of 27th ACM Conference on Human Factors in Computing Systems (CHI 2009), Boston, Mass, April 04-09, pp. 2051–2060. ACM, New York (2009)
11. Brezinka, V.: Treasure Hunt – a serious game to support psychotherapeutic treatment of children. In: Andersen, S.K., et al. (eds.) eHealth Beyond the Horizon – Get IT There, Studies in Health Technology and Informatics, vol. 136, pp. 71–76. IOS Press, Amsterdam (2008)
12. Rizzo, A., Klimchuk, D., Mitura, R., Bowerly, T., Buckwalter, G., Kerns, K., Randall, K., Adams, R., Finn, P., Tarnanas, I., Sirub, C., Ollendick, T., Yeh, S.: A Virtual Reality Scenario for All Seasons: The Virtual Classroom. In: The Proceedings of the 11th International Conference on Human Computer Interaction, Las Vegas (2005)
13. Harash, C.: Earthquake in Zipland. Earthquake in Zipland: Edutainment Game for Chirldren of Divorce (2010), `http://www.ziplandinteractive.com/` (called at January 2011)
14. Bonney, H.: Zappelix zaubert. PC-Lernspiel zum Training von Aufmerksamkeit und Impulskontrolle. MultiMediaManufaktur GmbH, Frankfurt (2010)
15. Olson, C.K.: Children's Motivations for Video Game Play in the Context of Normal Development. Review of General Psychology 14(2), 141–146 (2010)
16. Simms, D.C., O'Donnell, S., Molyneaux, H.: The use of virtual reality in the treatment of posttraumatic stress disorder (PTSD). In: Shumaker, R. (ed.) VMR 2009. LNCS, vol. 5622, pp. 615–624. Springer, Heidelberg (2009)
17. Rizzo, A., Graap, K., Mclay, R., Perlman, K., Rothbaum, B., Reger, G., Parsons, T., JoAnn, D., Pair, J.: Virtual Iraq: Initial Case Reports from a VR Exposure Therapy Application for Combat-Related Post Traumatic Stress Disorder. In: Virtual Rehabilitation 2007, pp. S.124–S.130 (2007)
18. Botella, C., Quero, S., Lasso de la Vega, N., Banos, R.M., Guillén, V., García-Palacios, A., Castilia, D.: Clinical issues in the application of virtual reality to the treatment of PTSD. In: Roy, M. (ed.) Novel approaches to the diagnosis and treatment of posttraumatic stress disorder. NATO Security Through Science Series, vol. 6. IOS Press, Amsterdam (2006)
19. Difede, J., Cukor, J., Patt, I., Giosan, C., Hoffmann, H.: The application of virtualreality to the treatment of PTSD following the WTC attack. Annals of the New York Academy of Sciences 1071, 500–501 (2006)
20. Kendall, P.C., Finch, A.J., Auerbach, S.M., Hooke, J.F., Mikulka, P.J.: The State-Trait Anxiety Inventory: A systematic evaluation. Journal of Consulting and Clinical Psychology 44, 406–412 (1976)

Virtual Reality as Knowledge Enhancement Tool for Musculoskeletal Pathology

Sophia Sakellariou[1], Vassilis Charissis[2], Stephen Grant[3], Janice Turner[4], Dianne Kelly[4], and Chistodoulos Christomanos[5]

[1] Gartnavel General Hospital
Department of Radiology, Glasgow, UK
[2] Glasgow Caledonian University, School of Engineering and Computing, Division of Computing and Creative Technologies, Glasgow, UK
[3] Ayr Hospital, Department of Trauma and Orthopaedics, Ayr, UK
[4] NHS Education for Scotland, Department of GP Education, Glasgow, UK
[5] Core Lab + Ltd., Software Developing Group, Glasgow, UK
s.sakellariou@nhs.net, v.charissis@gmail.com, stephen.grant1@nhs.net, Janice.Turner@nes.scot.nhs.uk, Diane.Kelly@nes.scot.nhs.uk, christomanos@gmail.com

Abstract. Contemporary requirements of medical explanatory resources have sparked the initiative of developing a unique pilot application which could use real-time 3D visualisation in order to inform General Practitioners (GPs) and allied health professionals as well as educate patients on musculoskeletal issues and particularly lower back pain. The proposed application offers a selection of 3D spinal anatomical and pathological models with embedded information. The interface elements adhered to previous studies' suggestions that the knowledge acquisition and ultimately the understanding of such motley three-dimensional subjects typically entails a strong grasp of the 3D anatomy to which it relates. The Human-Computer Interaction is simplified in order to empower the user to explore the healthy and pathogenic anatomy of the spine without the typical real-life constrains. The paper presents the design philosophy of the interface and the evaluation results from twenty user trials. Finally the paper discusses the results and offers a future plan of action.

Keywords: VR, 3D, HCI, Musculoskeletal, Medical Education, visual interface, Low Back Pain.

1 Introduction

Musculoskeletal (MSK) pathology and particularly lower back pain is an area of common interest for a large segment of medical practitioners including orthopedic surgeons, radiologists, general practitioners (GPs), physiotherapists and podiatrists amongst others [1]. Yet it is evident that the complexity of structures in the spine and the various pathologies that may appear are hindering significantly the learning process for all the aforementioned groups. The contemporary training and continuous professional development methods are unable to cover fully the intricate details of the pathologies occurring [2]. In particular the two-dimensional depictions or pseudo-3D

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 54–63, 2011.

images do not convey the appropriate information to the medical practitioners. To this end, previous studies demonstrated that the knowledge acquisition and ultimately the understanding of such motley three-dimensional subjects typically entail a strong grasp of the 3D anatomy to which it relates [3, 4].

Adhering to these requirements, we developed a pilot application, which uses a simplified interface in order to inform and further educate the interested parties. In particular the novel application offers a selection of real-time, stereoscopic, 3D spinal anatomical and pathological models with embedded information which can be accessed through the interaction with the 3D models. The development of the models preceded an exhaustive collection and analysis of CT scans, MRI scans and high-definition photographic data which informed the virtual reality characteristics of the model. The modeling process was further improved by constant consultation from the medical doctors which provided crucial clinical information. The complete highly detailed 3D models entail photo-realistic structures of the spinal vertebrae, the relevant ligaments, muscles, lymphatics, nerves and blood supply of that region.

Aiming for a user-friendly system the Human-Computer Interaction (HCI) is simplified in order to allow doctors of all grades to explore the healthy and pathological anatomy of the spine. The interface interlinks with a number of additional information complementing the real-time 3D data by high-definition (HD) explanatory animations, real-case scenarios, and problem –based learning (PBL) cases which lead the user through the presentation and reasoning of several lower-back pain case scenarios.

In order to evaluate this prototype system we invited twenty clinical users to assess the functionality and the context of the application. Additional information was gathered by sophisticated eye-tracking system which presented clear measurements regarding eye-accommodation, visual stimuli, and gaze-time during the trials. These preliminary user-trials offered constructive feedback and revealed great promise in the system with the derived results indicating better anatomy and pathology understanding. Interestingly the system enabled the doctors to accelerate the diagnostic triage process and create a 3D mental map of the context during the evaluation process. A number of issues, however, should be dealt with in the future stages of development, with particular interest on the Human Computer Interaction (HCI), which requires additional functionalities without spoiling the existing simplified interactivity. As such a simplified and customisable HCI could enable the user to interact with the system without the burden of unusable information or excessive text. To this end we envisage the development of the future version of the application which will enable the medical practitioners to customise their own systems in accordance to their interests and their expertise. Notably this could potentially provide both an aid to the understanding of the detailed regional anatomy for doctors in training as well as a highly interactive platform of training for those interested in surgical interventions. Finally the subjective feedback from the evaluation drew attention to the potential use of an even more simplified system which could form part of educating patients of their condition and obtaining informed consent.

Overall this paper discusses the challenges involved in the development process of the Virtual-Reality learning application and the HCI design developed for this system. Furthermore this work introduces the visual components of the interface and presents the outcome of a preliminary evaluation. The paper concludes with a future plan of of work which aims to expand the context and interactivity of the system so as to enable other types of users (i.e. undergraduate/postgraduate medical trainees and patients) to access meaningful information without the time and physical constrains of laboratories. Finally this prototype application entertains the use of virtual reality for training purposes of General Practitioner doctors and offering the same user-experience through a web-based version of this application.

2 Contemporary Training Issues

Anatomy training, in the UK in particular, has presented medical educators with a plethora of challenges in recent years as the undergraduate educational focus is shifted away of traditional theoretical pure knowledge acquisition and targeted towards interactive experiential learning. Cadaveric opportunities of anatomical knowledge acquisition are becoming scarce due to a shortage of cadaveric material and limiting practical constrains associated with such endeavours. This situation, in combination with the shortening of the post graduation training years with the introduction of the fast track structured specialty training has limited exposure of trainees to appropriate practical anatomy learning opportunities and has resulted in a large proportion of the current medical and allied health trainees recognising a pressing need for alternative anatomy and pathology experiential based interactive educational packages.

Furthermore, the introduction of the good medical practise guidelines from the GMC and associated publications from the other professional bodies has placed an increased emphasis on continuous education and re-validation [5, 6, 7]. Continuous professional development applications can provide crucial assistance in meeting the clinicians' educational demands. A vast variety of paper based and digital packages have already been developed to this end, however, the end users have expressed concerns regarding their inaccessibility, associated time and space constrains as well as the lack of interactivity.

With particular reference to the anatomy and pathology of back pain, it was found that most of the available educational resources was in either actual 3D plastic model form or in the form of traditional text and 2D illustrations. Large datasets of data capable of 3D reconstruction (such as patient CTs and MRIs) were also available on request for further education, however, these required expert input by a trained specialist. Within a community, rather than hospital, setting, where most of the general practitioners as well as most of the allied professionals involved in the study work, realistically, the only educational avenues available were either in the form of traditional paper based educational material or material available through alternative digital avenues which could potentially be transferable through the world wide web.

Back pain was chosen as the pilot study case as it is one of the most common complaints expressed by patients in GP consultations in the UK. It makes up 30% of out-patient physiotherapy work in Scotland, translating to about 12000 new referrals per annum [8]. Although most of this work (~85%) relates to chronic back pain, however, there are a few pathologies disguising as back pain, which are far more serious and should not be missed. A national physiotherapy low back pain audit[8] identified a gap in stuff education and recommended the development of a web based easily accessible package that would raise awareness and understanding of acute back pain and its management [8].

3 Visualisation Methods

A plethora of digital applications and visualisation methods have been recently developed in order to educate different professionals and especially the medical practitioners, aiming to decrease the learning curve and act as a knowledge enhancement tool. Noticeably a number of aesthetically pleasing visualisations have been produced in a compulsive and fractious manner.

Two dimensional representations and alphanumeric diagrams of complex anatomical data have failed evidently in the past to convey knowledge, mainly due to the fact that the human body itself and the environment that surrounds it, are by definition three-dimensional. As such it is impossible to depict the three-dimensional complexity of multiple layers of the human body through two-dimensional illustrations [9].

In turn the digital formats appeared largely through the pseudo-3D visual representations which offered a different learning avenue than the traditional methods. Although this concurrent approach is considered a breakthrough, the limitations of such systems are constraining the user from developing their own learning pattern. The pseudo-3D models are merely a collection of still images produced from an object through a 360 degrees rotation. The pre-set nature of these models forces the user to add or remove layers of a model in an un-involving rotational manner which does not encourage the investigation of the object from different angles or the non-uniformly application of cross-sections.

Real-time 3D visualisation unlocks this interactivity as it allows the user to investigate the 3D model and interact with it in an infinite number of ways. The manipulation of the presented data is not constrained by pre-set factors but only from the user's motivation to interact with the system. Furthermore, the interaction tools that could be developed for such application can be customised depending on the learning outcome and extend users' experience beyond the typical axis rotation of images or the aloof viewing of multiple animations.

4 VR Interface

Recent studies have advocated that 3D and particularly Virtual Reality (VR) training methods abbreviate significantly the learning curve [10]. Our previous experimentation with different VR and Augmented Reality (AR) systems offered very similar results to the aforementioned [3, 4, 11]. Based on our prior experience we aimed to develop an interface that could accommodate the different visual and auditory information that the particular group requested (GPs).

Yet being aware of the potential system limitations as well as the different levels of computer literacy amongst users we strived towards a simple interface which could compliment the real-time 3D application with a variety of specialised still images, explanatory animations and related documentation that could assist and enhance the knowledge of the specific users [12, 13].

The interface design followed an organic method for distributing information which were interlinked though different connections. As such a single level of the interface could be approached by a number of different points inside the system without constraining the user to follow a specific root. The focus point of the system was evidently the real-time 3D models which can be viewed either on mono or stereo depending the user's equipment as presented in Figure 1.

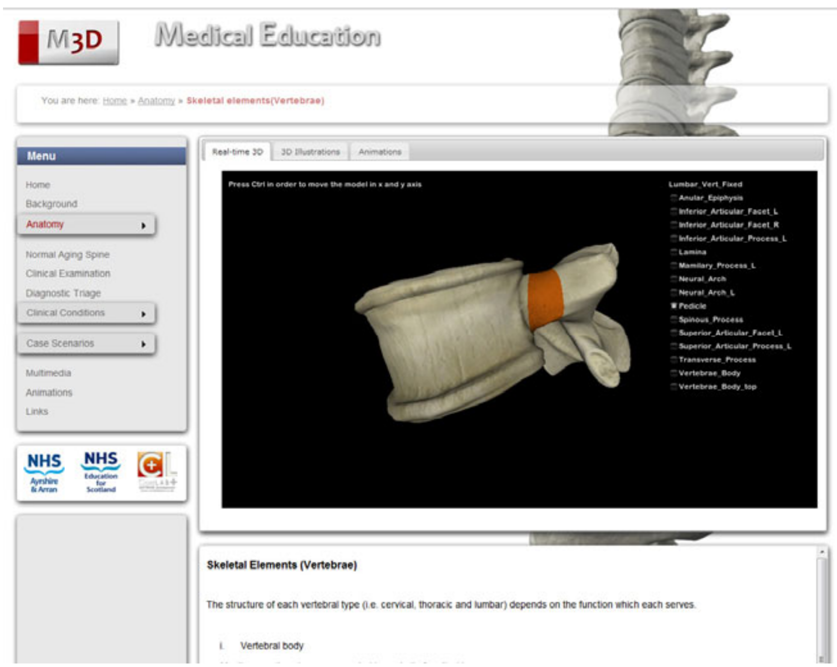


Fig. 1. Screenshot of the main application interface

Yet adhering to the simplistic approach of information conveyance the stereo to mono alteration can be achieved automatically as soon as the system recognises the specific equipment.

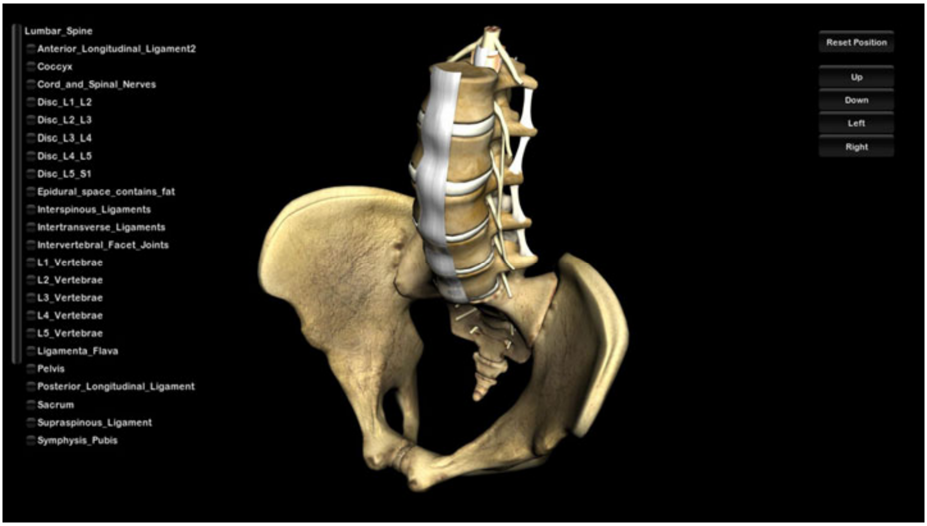


Fig. 2. Screenshot of the real-time 3D model of the lumbar spine and pelvis in full-screen

The interactivity for each model was kept to a minimum at this stage offering to the user the flexibility to rotate, zoom, and pan the model. Additionally all the requested sections of each model could be identified and the relevant terminology is presented on the side of the screen in a semi-transparent layer which aims to avoid any unesseccery visual conflict between the main visual information (i.e. 3D model) and the supplementary information (i.e. alphanumeric data).

The system also enables the user to maximise in full-screen the real-time 3D model and maintain only the related terminology (Figure 2). Currently we are in process of developing a new range of interaction tools which will improve significantly the user experience and learning process. This group of tools will empower the user to dissect in non-uniformly manner the model, intersect CTs and MRIs related to the model, apply selective transparency, move individual parts and compare one or more models in the same window. The latter could be beneficial for training regarding identification of similar conditions, yet in different stages (i.e. malignant cord compression).

5 Context Development

A photorealistic, real-time 3D model of the spine was constructed drawing on data from traditional illustrations, large CT and MRI datasets and under guidance by specialist medical practitioners. This was realistically textured and annotated via an interactive system whereby selecting a label would indicate on the model the exact position of the labelled structure and vice versa. Detailed, but not exhaustive anatomy labelling was included in the 3D model. Structures trivial to general medical anatomy and pathology were deliberately omitted. On the other hand, structures commonly affected by pathology were highlighted in case studies. The visual input of information was further enriched by greatly detailed still images and high-definition animations.

Five different aetiologies of acute lower back pain were illustrated using the 3D models, animations, photorealistic still images and correlating with medical diagnostic imaging. Thus mechanical back pain, cauda equina syndrome, spinal stenosis, ankylosing spondylitis and malignant infiltration of the spinal cord were presented as medical case studies with subsequent questions on diagnosis and management.

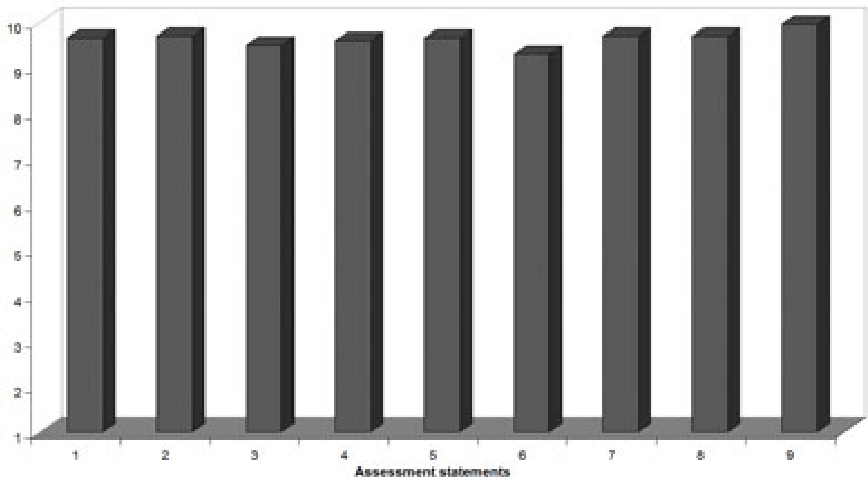
An extended optimisation of the 3D models was further applied in order to improve their refresh rate when presented in a 3D stereoscopic format. Notably the optimisation process was applied in a two fold approach, firstly to the actual geometry and in turn to the textures and shaders. The final optimisation outcome enabled the models to be viewed by minimum specification computers and handheld mobile devices.

6 Evaluation

Preliminary evaluation of the system was undertaken through interactive sessions of the end users with the 3D real-time environment and subsequent anonymised scoring of the user experience and the medical context. Comments relating to the educational value of the associated factual information of the proposed application are beyond the scope of this paper and will not be further analysed here.

Evaluation of the interface was two fold: Subjective user evaluations were gathered regarding the usability, interactivity and usefulness of the interface and the use of the real time 3D models, as scores on a Likert scale. Objective evaluation was also performed by a neutral observer on the usability of the system as evidenced by the user/3D model interaction, time to retrieval of information and overall willingness to explore the application. Eye tracking equipment also monitored the user's focusing points and indicated whether the interface augmented or hindered the learning process.

Quantitative and qualitative analysis of the Likert questionnaire and of free text comments and suggestions was performed (Figure 3). Twenty users, 60% general practitioners and 40% physiotherapists accessed and evaluated the application. Their prior familiarity with 3D technology ranged from none to moderate with only 5% having prior exposure to 3D within the context of educational material. The users' IT competence also ranged significantly with the majority (80%) subjectively and objectively scoring average IT competence. 10% subjectively scored themselves as proficient, but objectively only 5% was. 5% (User 8) scored themselves as moderately competent but on objective assessment was of very limited IT competence, a factor that affected the user's interaction with the application and the subsequent user's assessment. Despite of this limitation of this user, the scores were included in the analysis. Overall, end users were overwhelmingly satisfied with the interactivity provided by the 3D model and found information portrayed sufficient and easy to access. It is interesting to note that 95% of the users strongly agreed that the application can be used as an educational tool for both practitioner and patient and that they would use it as an aid to teaching. Even the user who found interaction with the system difficult scored it relatively high for conveying information for teaching purposes.



1	The 3D real time models are useful for understanding MSK anatomy
2	The 3D real time models are useful for understanding MSK pathology
3	The 3D real time models are useful for explaining to the patient MSK anatomy
4	The 3D real time models are useful for explaining to the patient MSK pathology
5	I can manipulate easily the 3Dreal time models and interact with them
6	I can easily correlate the 3Dmodel sections to their names in the list provided
7	The 3D models provide sufficient structural detail
8	The 3D models are realistically textured
9	I would use the real time 3D models for teaching purposes

Fig. 3. Likert scores, 3D model evaluation
(1-10 scale, 1=Strongly disagree, 10=Strongly agree).

Some of the collated comments that predominated in our qualitative analysis:

- 90% of users – Realistic, accurate, interactive 3D facilitate learning.
- 85% of users – Easy to use/navigate
- 80% of users – Excellent integration of visual stimuli (3D model) and factual information (text/labels)
- 95% of users – Asked for expansion of the application to cover other body areas/pathologies
- 60% of users - Suggested reduction in the volume of the associated factual information and indicated that the 3D models and animations conveyed the required information more effectively.

Extracts from individual users' comments:

- "User friendly – things are in common sense areas"
- "Use of 3D is intuitive. Super imposed pathologies and medical scans on 3D models facilitate learning by demonstrating real life pathology and engaging the user"
- "Could easily be tailored to patient education, facilitating patient enablement"

The aforementioned feedback and observations were further verified by the preliminary analysis of the eye-tracking generated data which demonstrated an approximately 90% concentration of eye gaze on the 3D models at in any given page of the interface as illustrated in the Figure 3 below.

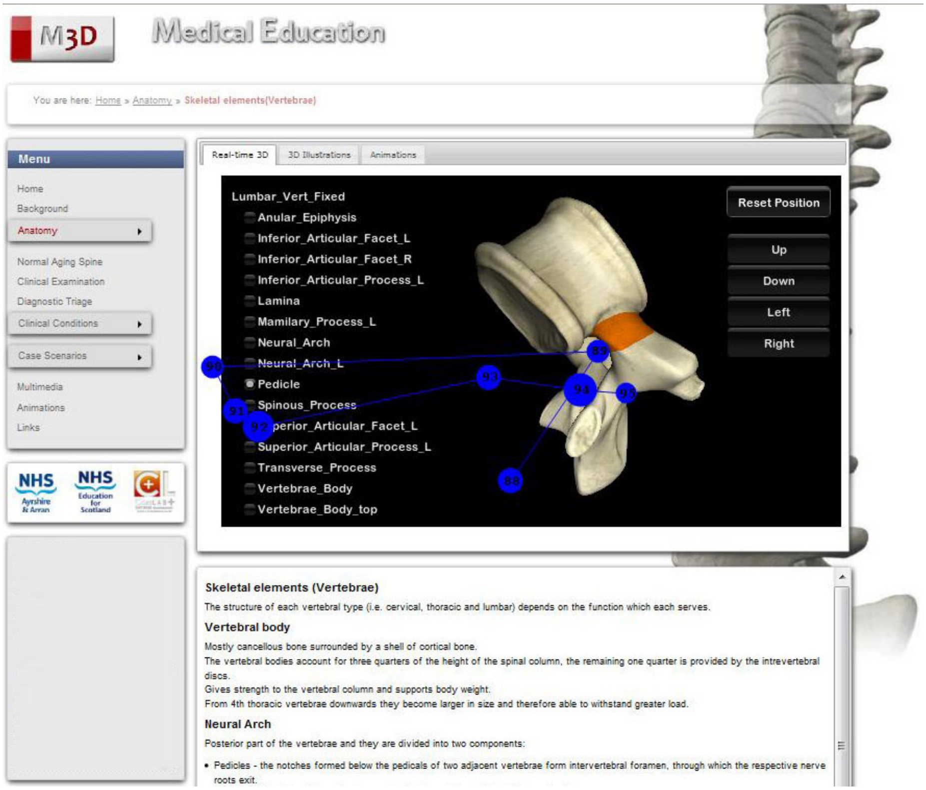


Fig. 4. Screen capture of the recorded video footage presented the eye-tracking results

7 Conclusions

Overall this paper discusses the challenges involved in the development process of the Virtual-Reality learning application and the HCI design developed for this system. Furthermore this work introduces the visual components of the interface and presents the outcome of a preliminary evaluation. The proposed interface has been evaluated with the use of both quantitative and qualitative methodologies. The derived results are promising as the vast majority of the users enjoyed the experience and refreshed or accumulated the indented material by exploring the 3D real-time data in a simplified, timely manner.

Our tentative plan of future work aims to expand the context and interactivity of the system so as to enable other types of users (i.e. undergraduate/postgraduate medical trainees and patients) to access meaningful information without the time and physical constrains of laboratories. Additionally aspiring to increase the accessibility of the system the prototype VR application entertains the use of virtual reality for training purposes of General Practitioner doctors and offering the same user-experience through a web-based version of this application which currently can be visited on the www.3dmedicaleducation.co.uk. Consequently we plan a second series of system trials which will further inform our thinking and design of interface tools.

References

1. Turney, B.W.: Anatomy in a Modern Medical Curriculum. Annals of the Royal College of Surgeons of England 89(2), 104–107 (2007)
2. Lu, Q., Luo, S.: Primary Research of Digital Atlas of Human Anatomy on Virtual Reality. In: Proceedinggs of the 2nd International Conference on Bioinformatics and Biomedical Engineering (ICBBE 2008), Shanghai, China, pp. 2442–2445 (2008)
3. Charissis, V., Zimmer, C.R., Sakellariou, S., Chan, W.: Development of Virtual Reality Simulation for Regional Anaesthesia Training. In: Proceedings of the International Annual Symposium of IS&T/SPIE, The Engineering Reality of Virtual Reality, San Jose, California, USA, January 17-21 (2010)
4. Sakellariou, S., Ward, B.M., Charissis, V., Chanock, D., Anderson, P.: Design and Implementation of Augmented Reality Environment for Complex Anatomy Training: Inguinal Canal Case Study. In: Procceedings of the Human Computer Interaction International 2009, San Diego, California, USA (2009)
5. Chartered Society of Physiotherapy, Clinical Guidelines for the Physiotherapy Management of Persistent Low Back Pain (LBP): Part 1 excersise (2006)
6. Chartered Society of Physiotherapy, Clinical Guidelines for the Physiotherapy Management of Persistent Low Back Pain (LBP): Part 1 excersise (2006)
7. Royal College of General Practitioners (RCGP), Clinical Guidelines for the Management of Acute Low Back Pain (1999)
8. NHS Quality Improvement Scotland, National Physiotherapy Low Back Pain Audit, Improving Back Care in Scotland (2009)
9. Garg, A., Norman, G., Sperotable, L.: How medical students learn spatial anatomy. In: The Lancet 2001, pp. 363–364. Elsevier, Amsterdam (2001)
10. Sang-Hack, J., Bajcsy, R.: Learning Physical Activities in Immersive Virtual Environments. In: Fourth IEEE International Conference on Computer Vision Systems (ICVS 2006), p. 5(2006)
11. Ward, B.M., Charissis, V., Young, I., Rowley, D.I., Anderson, P.: Can Virtual Reality Augment Postgraduate Anatomy Teaching? Surgical Education: A Brief Enquiry into Applied Virtual Anatomy 1, 1 (2008)
12. Bay, B.H., Ling, E.A.: Teaching of Anatomy in the new millennium 48(3), 182–183 (2007)
13. Chittaro, L., Ranon, R.: Computers & Education. Web3D technologies in learning, education and training: Motivations, issues, opportunities 49(1), 3–18 (2007)

Study of Optimal Behavior in Complex Virtual Training Systems

Jose San Martin

Universidad Rey Juan Carlos, Madrid
jose.sanmartin@urjc.es

Abstract. In previous works we have studied the behavior of simple training systems integrated by a haptic device basing on criteria derived from Manipulability concept. The study of complex systems needs to re-define the criteria of optimal design for these systems. It is necessary to analyze how the workspace of two different haptics, simultaneously on the same model, limits the movement of each other. Results of the new proposed measures are used on Insight ARTHRO VR training system. The Minimally Invasive Surgery (MIS) techniques use miniature cameras with microscopes, fiber-optic flashlights and high definition monitors. The camera and the instruments are inserted through small incisions on the skin called portals. The trainer uses two PHANToM OMNi haptic devices, one representing the camera and other the surgical instrumental.

Keywords: Haptics, Workspace Interference, Manipulability, Optimal Designing.

1 Introduction

When developing a VR simulation, on the one hand we are moving in the real world by manipulating a haptic device; this movement corresponds to a movement of an object in our simulation, that is, in the virtual environment. So we must first study the portion of space that can be achieved with our device, called the real workspace (RW). Therefore should be made according to his characterization of an efficiency measure based on the criterion of Manipulability [1], which will allow us to redefine the subareas we want to work from all the available area. So, according to the virtual environment to define, called virtual workspace (VW), this portion of the RW must be chosen.

The Minimally Invasive Surgery (MIS) techniques use miniature cameras with microscopes, tiny fiber-optic flashlights and high definition monitors. The camera and the instruments are inserted through small incisions on the skin. The visual control of organs and surgical instruments is done thanks to the image provided by the video camera [2][3]. These procedures are complex to perform and require specialized training in order to obtain proficiency in the skills [4]. Therefore this paper proposes a methodology to optimize the design of these training systems, so that depending on the surgical technique to implement, it is possible to define an optimal configuration of virtual training system for each case.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 64–72, 2011.

2 Manipulability Solid

Among several indices that allow us to study the quality of systems involving several haptic devices we have chosen the well known concept of Manipulability [5] [6] [7], since it allows a quantitative value to the quality of the system.

Manipulability of a device is its ability to move freely in all directions into the workspace [8]. In these terms the Manipulability for each device configuration, is a tool for evaluating the quality and the performance in the designing of a manipulator device [1] [9] [10].

From the original formulation of Yoshikawa, we use the formulation of Manipulability proposed by [11]:

$$\mu = \sigma_{min} (J_u)/\sigma_{max} (J_u) \tag{1}$$

Where:

σ_{min} and σ_{max} are the minimum and maximum singular values of J_u, upper half of the manipulator Jacobian matrix.

For each haptic we can define a volume as the Manipulability Solid associated to the device (fig. 1).

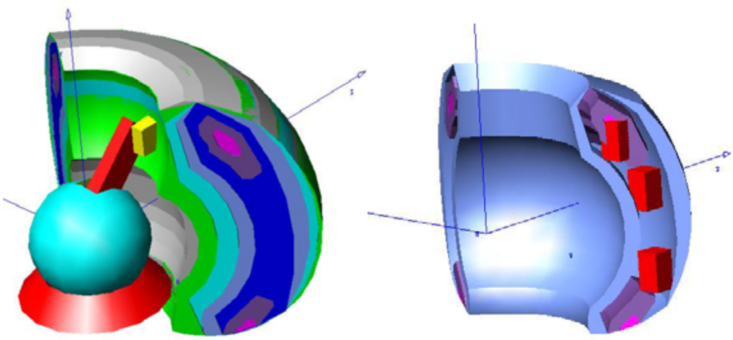


Fig. 1. 3D Solid of Manipulability (left). Detail of positioning options of a VW inside the RW (right).

From whole the RW a portion is selected: the virtual environment which we called Virtual Workspace (VW), in red in fig. 1-right. We place VW inside RW taking advantage of best zones according to Manipulability index (different options of positioning a VW are presented).

3 Useful Manipulability

During a simulation session we find that there are zones of VW where the specific task is effectuated. So it is desirable than the haptic device provides its best performance in these zones. For a specific task to be done, we define Navigation Frequency Map-NFM as the VW where we assign to each voxel the number of times is visited when doing this task.

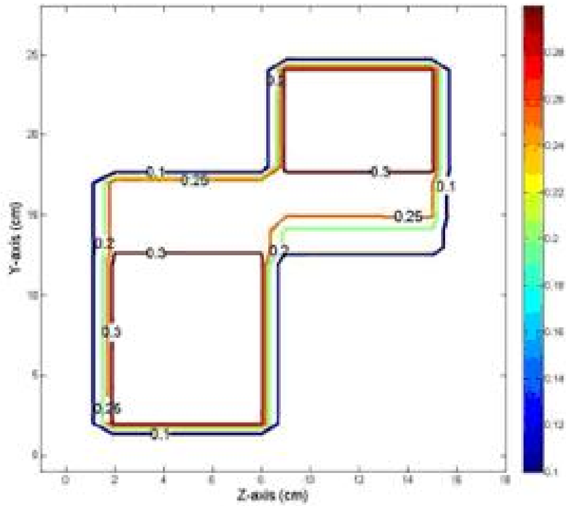


Fig. 2. 2D Portion of NFM-Diagnostic Arthroscopy. NFM is a 3D object in XYZ, figure represents a portion for an X defined value. Level curves represent frequency values in 10e-3.

We discretize RW in a grid of voxels so we call μ_{ijk} the value of manipulability calculated using (1) in each cell (i, j, k).

$$\hat{\mu}_v = \sum_{ijk} \mu_{vijk} \cdot f_{ijk} \tag{2}$$

where:

$\mu_{v\,ijk}$ is the distribution of Manipulability into a cell.

f_{ijk} is the frequency of visits sampled during a simulation session in a cell.

Figure 2 shows an example of a NFM obtained from the simulation of a Diagnostic Arthroscopy.

4 A New Algorithm for Optimal Positioning

The design of the proposed system requires positioning the VW inside the solid of Manipulability. This is the intersection of two solids, the bigger, the RW that spans the space where we can work on the other side the VW, which corresponds to the RW subsection we want to use.

This intersection affect zones with different values of μ (different colored volumes v_i in figure 1). The use of this concept has the difficulty of finding the best relative positioning, i.e. what is the best area of RW for that size of VW, and for the application which we will do in this virtual environment

We modify the previously used global optimization algorithm for solving it. The main advantage obtained with the use of this modified algorithm compared to the previous, Simulated Annealing-SA [12], is that in the case of the use of SA the ability to converge sometimes was compromised if there was much difference between the size of RW and VW, so we could stay in local minima. Now the obtained search process is therefore more general.

4.1 Interference Factor-IF

Optimal navigation of a single device has been defined in previous work by the coefficient of Useful manipulability [13]. However, when the system consists of two devices, the behavior of one influences the other. While the two haptic devices do not perform a Dual Arm Cooperative Task Execution [14], there is interference because we have two haptics working in the same VW. We define Interference Factor (IF) as the index to quantify that effect. For its definition is necessary to study NFM tasks of each device, as well as introducing a time base. So taking into account not only how many times you visit a voxel of VW, but in what instant, and the proximity of the voxels that both Haptics visit in the same moment.

Additionally we must consider that there is a task, in this case the tool's operation, which can be primary. In this case we can think about minimizing the interference factor of the camera on the tool, providing NFM modification of the OMNi camera, looking for a new optimal path.

5 Results

In Insight ARTHRO VR, containing two devices, the application is centered on a virtual model of the human shoulder. We compare the optimal XYZ relative positioning for a set of different operations. We obtain these optimal positioning and the $\hat{\mu}_v$ values corrected with the new IF index.

Interference factor is calculated considering a time base. To do this, when we create the NFM, is incorporated in each voxel, not only the value of visits to this voxel, but also the instant of time of each visit. So when we study the comparison of dual work with two haptic devices, we must consider where to find the End Effector (EE) of each device at all times.

The IF is a function of the distance between the EE of each device so that if the distance is reduced, should be incorporated a real interference in the devices movements, since the rate of Manipulability of individual haptic, is going to be conditioned by the movement of the other.

The IF is modeled as an exponential value and corresponding to a maximum value of 1, without interference from a distance that our case is 5 voxels. For the positioning based in Simulated Annealing + IF a cell for the size of VW of 2mm is selected. Then IF is effective in a distance of 10mm between both EE. Moreover, with this size of the grid we can simplify assuming $\mu_{v\,ijk} = \mu_{ijk}$ calculated by (1), that is, we can suppose in each cell the value of the index of Manipulability is constant, so (2) can be simplified to:

$$\hat{\mu}_v = \sum_{ijk} \mu_{ijk} \cdot f_{ijk} \tag{3}$$

We have to realize the study of a system that is going to involve two OMNi devices, provided that in case of study, an operation of minimally invasive surgery, they are going to play roles of video camera and operation tool. Therefore there have been defined the corresponding VW of VW-camera and VW-tool. In these VW both values of Navigation Frequency Map (NFM) (Figure 2) must be indicated. With this information we incorporate the value of time each voxel is visited and the instant of the visit.

Once obtained the maps of frequencies for each of two devices there begins the process of search of the ideal positioning. The initial parameters of the Simulated Annealing [15] are:

Initial Error: error_i = 0.12
Initial Probability: pr_i = 0.28
Initial Temperature: $temp\_i = \frac{-error\_i}{\log(pr\_i)} = 0.125$

The "condition of cooling " is temp=temp*0.92. The process will continue while the condition is fulfilled of temp > 0.005.

The initial placement of the VW must be arbitrary in order to avoid local minima. As the temperature is modified, jumps of position are going to be realized, and we have introduced too a bigger jumps in this version. In every temperature the variation of position for each of the directions XYZ is calculated by the formulae:

$$\partial x = \frac{k}{6.5} \cdot random \cdot \frac{temp}{temp\_i} \cdot \quad \partial y = \frac{k}{6.5} \cdot random \cdot \frac{temp}{temp\_i} \cdot \quad \partial z = \frac{k}{6.5} \cdot random \cdot \frac{temp}{temp\_i} \cdot$$

where:

k is a constant proportional to the resolution of the grid
random is a random Lumber between 0 and 1.

The conditions of acceptance are one of the following:

$$\hat{\mu}_{\text{v}} > \hat{\mu}_{\text{v-accepted}} \quad p_r > p_a \; .$$

being:

pr – value of the probability at a temperature: $p_r = \exp\left(-\frac{\hat{\mu}_{\text{v-accepted}} - \hat{\mu}_{\text{v}}}{temp} \right)$
pa – a random number between 0 and 1.

If this positioning is accepted, this cell turns into the new ideal point to place the VW:

$$\hat{\mu}_{\text{v-accepted}} = \hat{\mu}_{\text{v}}$$

However, in each case, before considering the value $\hat{\mu}_{\text{v-accepted}}$ is necessary to check whether there have been problems of interference with this configuration. IF factor is a characteristic of the operation being simulated, so there are interventions within the same virtual environment, where interference will occur and other interventions that do not.

Thus, the optimal solution for each of two devices OMNi will come indicated by the position XYZ that corresponds to this ideal value of convergence. The results indicate the position XYZ where the optimal performance for the OMNi is obtained from the coordinate origin at the center of OMNi device, and the corresponding value of $\hat{\mu}_v$.

5.1 Different Virtual Scenarios

In order to study the effect of modifying NFM with the same haptic device in the same virtual environment, we selected additional surgeries techniques:

Surgery for shoulder instability. The Bankart lesion is a tear on the labrum in the lower part of the shoulder joint [16]. A SLAP lesion involves the labrum and the ligament on the top part of the shoulder joint [17]. The first operation was a shoulder arthroscopy where most part of the work is made in the border of the glenohumeral zone. The problem to solve is called **Bankart lesion**, and the intervention consists of several steps: Bankart's injury is opened with arthroscopic instrumentation to obtain osseous bleeding tissue. After this, mini-screws have to be placed at the edge of the glenoid cavity and threads of suture across the labrum and from the glenohumeral ligaments. Then the ligaments and the labrum must be knotted and prefixed to the edge of the glenoid cavity. Finally the opening is closed. According to the analysis of the movements of the surgeon in this type of intervention, we created a NFM (Figure 3-1). The **SLAP lesion** involves the work in the top part of the joint, so both of them Bankart and SLAP are located in the shoulder joint but on the bottom or in the top. For SLAP lesion we created a NFM (Figure 3-2).

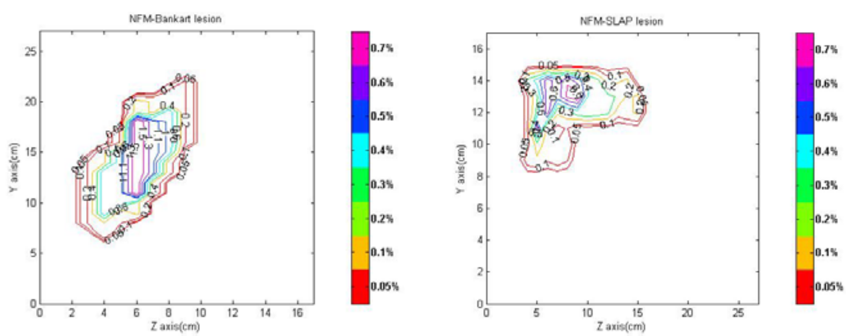


Fig. 3. 3-1. Frequency map corresponding with the Bankart lesion. 3-2. Frequency map corresponding with the SLAP lesion.

Surgery for impingement syndrome. Damaged or inflamed tissue is cleaned out in the area above the shoulder joint itself. The surgeon may also cut a specific ligament and shave off the under part of a bone. This under part of the bone is called the acromion. Removing a part of the acromion can stop the tendons of the rotator cuff from rubbing on the bone. This type of surgery is called an **Acromioplasty** [18]. So, in this intervention, most part of the work is done in the surroundings of the acromion. For Acromioplasty we created a NFM (Figure 4).

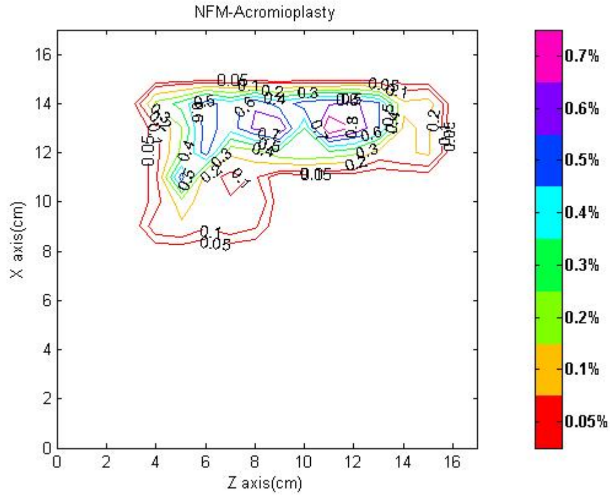


Fig. 4. Frequency map corresponding with Acromioplasty

Rotator cuff disorders. These are irritations in or damage to tendons around the shoulder. The selected operation is called Bursectomy. It is a shoulder arthroscopy where most part of the work is made in the subacromial area. The subacromial bursa is a tissue that serves as shock absorber and that diminish the rubbing between the acromion and the supraspinous ligament [19]. The patient with problems of the muff has an inflammation and hypertrophy of the bursa, called bursitis. The Bursectomy consists of drying out up the inflamed bursa. See the corresponding NFM in Figure 5.

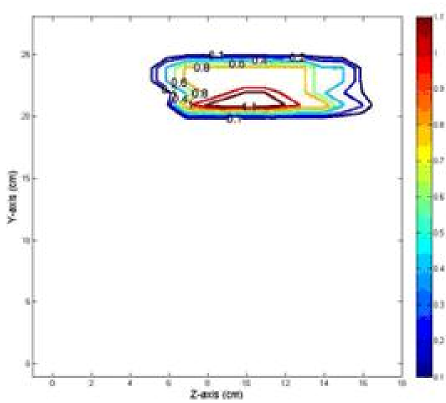


Fig. 5. Frequency map corresponding with Bursectomy

The results obtained by high resolution whole-exploration including IF factor, for optimal position XYZ are:

Bankart lesion:

OMNi-tool: $\hat{\mu}_V = 0.9176$, position (-41, 139, 252) millimeters.

OMNi-camera: $\hat{\mu}_V = 0.897$, position (-10, 112, 109) millimeters.

SLAP lesion:

OMNi-tool: $\hat{\mu}_V = 0.9384$, position (-38, 167, 263) millimeters.

OMNi-camera: $\hat{\mu}_V = 0.897$, position (-9, 98, 113) millimeters.

Acromioplasty:

OMNi-tool: $\hat{\mu}_V = 0.9384$, position (-40, 143, 234) millimeters.

OMNi-camera: $\hat{\mu}_V = 0.897$, position (-10, 107, 101) millimeters.

Bursectomy operation:

OMNi-tool $\hat{\mu}_V = 0.908$, position (51, 139, 250) millimeters.

OMNi-camera: $\hat{\mu}_V = 0.897$, position (-14, 115, 121) millimeters.

Acknowledgments. I would like to thank the anonymous reviewers for their helpful comments, Gracian Trivino and the rest of the GMRV group at URJC. This work has been partially funded by the Spanish Ministry of Education and Science IN2007-67188 and Cajal Blue Brain project.

References

1. Yoshikawa, T.: Manipulability and redundancy control of robotic mechanisms. In: Proceedings of IEEE International Conference on Robotics and Automation, vol. 2, pp. 1004–1009 (March 1985)
2. Agha, R., Muir, G.: Does laparoscopic surgery spell the end of the open surgeon? J. R. Soc. Med. 96, 544–546 (2003)
3. Ahlberg, G., Heikkinen, T., Iselius, L., Leijonmarck, C.-E., Rutqvist, J., Arvidsson, D.: Does training in a virtual reality simulator improve surgical performance? Surg. Endosc. 16, 126–129 (2002)
4. Moore, M.J., Bennett, C.L.: The Learning Curve for Laparoscopic Cholecystectomy. The American Journal of Surgery 170(1), 55–59(5) (1995)
5. Yokokohji, Y., San Martin, J., Fujiwara, M.: Dynamic Manipulability of Multifingered Grasping. IEEE Transactions on Robotics 25(4), 947–954 (2009)
6. Gallina, P., Rosati, G.: Manipulability of a planar wire driven haptic device. Mechanism and Machine Theory 37(2), 215–228 (2002)
7. Cho, C., Kim, M., Song, J.: Direct Control of a Passive Haptic Based on Passive Force Manipulability Ellipsoid Analysis. International Journal of Control, Automation, and Systems 2(2), 238–246 (2004)
8. Murray, R.M., Li, Z., Sastry, S.S.: A mathematical introduction to robotic manipulation. CRC Press, Inc., Boca Raton (1994)

9. Yamamoto, Y., Yun, X.: Unified analysis on mobility and manipulability of mobile manipulators. In: Proceedings. 1999 IEEE International Conference on Robotics and Automation, Detroit, vol. 2, pp. 1200–1206 (1999)
10. Yokokohji, Y., Yoshikawa, T.: Guide of master arms considering operator dynamics. Journal of dynamic systems, measurement, and control 115(2A), 253–260 (1993)
11. Cavusoglu, M.C., Feygin, D., Tendick, F.: A Critical Study of the Mechanical and Electrical Properties of the PHANToM Haptic Interface and Improvements for High Performance Control. Teleoperators and Virtual Environments 11(6), 555–568 (2002)
12. Weise, T.: Global Optimization Algorithms. Theory and Application. Online e-book under GNU Free Documentation License (2009), `http://www.it-weise.de/projects/book.pdf`
13. San Martin, J., Trivino, G.: Mechanical design of a minimal invasive surgery trainer using the manipulability as measure of optimization. In: Proc. IEEE Int. Conf. Mechatronics (ICM), Kumamoto, Japan, pp. 1–5 (May 2007)
14. Lee, S., Lee, J.M.: Task-oriented dual-arm manipulability and its application to configuration optimization. In: Proceedings of the 27th IEEE Conference on Decision and Control 1998, December 7-9, pp. 2253–2260 (1988)
15. Aragon, C.R., Johnson, D.S., McGeoch, L.A., Shevon, C.: Optimization by Simulated Annealing: An Experimental Evaluation Part II, Graph Coloring and Number Partitioning. In: Operations Research, vol. 39(3), pp. 378–406 (1991)
16. Phillips, B.B.: Arthroscopy of the upper extremity. In: Canale, S.T., Beatty, J.H. (eds.) Campbell's Operative Orthopaedics, 11th edn., ch.49. Mosby Elsevier, Philadelphia (2007)
17. Snyder, S.J., Karzel, R.P., Pizzo, W.D., Ferkel, R.D., Friedman, M.J.: SLAP lesions of the shoulder. Arthroscopy: The Journal of Arthroscopic & Related Surgery 6(4), 274–279 (1990) ISSN 0749-8063
18. Beasley Vidal, L.S., et al.: Shoulder injuries. In: McMahon, P.J. (ed.) Current Diagnosis and Treatment in Sports Medicine, pp. 118–145. McGraw-Hill, New York (2007)
19. Phillips, B.B.: Recurrent dislocations. In: Canale, S.T., Beatty, J.H. (eds.) Campbell's Operative Orthopaedics, 11th edn., ch. 45. Mosby Elsevier, Philadelphia (2007)

Farming Education: A Case for Social Games in Learning

Peter Smith and Alicia Sanchez

ADL Co-Lab, Defense Acquisition University
13501 Ingenuty Dr. Orlando Fl, 32826
Peter.Smith.CTR@adlnet.gov

Abstract. Social games have skyrocketed in popularity; much to the surprise of many in the game development community. By reinforcing individualized internalization of concepts while framing those experiences in terms of social activities, social games are filling a void not adequately filled by other games and may turn out to be power learning tools. Their potential use in education is still in its infancy as many consider how the characteristics unique to social games could be used within a learning paradigm. By creating asynchronous multiplayer environments and play dynamics designed to leverage both individual and collaborative goals, social games may foster long distance relationships and encourages reflection of the tasks preformed.

Keywords: Social Games, Social Networks, Learning Games, Serious Games.

1 Introduction

Social activities continue to evolve as technology moves more of society's activities from the real space to the virtual. Casual occurrences such as keeping in touch with friends and family, announcing accomplishments, sharing photos, or wishing someone happy birthday have in many cases become an online activity. One researcher, Valentina Rao suggests that social networks are a "Third Place" [1]. That is, an area dedicated to playfulness. Ray Oldenburg first introduced "Third Places" as a place that is separate from home or work [2]. In a "Third Place" people congregate to play, interact, and engage in discourse among other things that are generally playful in nature.

Rao suggests that online social networks have already begun to replace conventional "Third Places", removing the need for people to be in the same geographic location and moving that experience to a variety of virtual places. With the removal of geographical limitations comes the increase in both size and relevance of social networks. These virtual social activities and their ability to expand an individual's social network have begun to transcend the barrier of the "Third Space" making them integral at both work and home. Incorporating education and training into these "Third Places" might be an effective way to keep learners connected, to keep them motivated and to keep them learning.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 73–79, 2011.

1.1 Pervasiveness of Social Networks

Facebook is the undisputed heavy weight of the social networking world. In February 2011 they had over 500 million users who spend over 700 billion minutes on Facebook per month [3]. What is really impressive is the active users create an average of 90 pieces of content a month, resulting in 3.5 billion pieces of content per week. Image uploads alone equal 2.5 billion each month [4]. Although, other popular social media sites lag behind Facebook they are by no means small. Most of these sites tend to target specific demographic markets, and provide valued services for their audience. For example, LinkedIn targets professionals with services for advertising and looking for employment. While maintaining focused on their market they have managed an impressive 15 million monthly users and over 67 million total [5].

1.2 Pervasiveness of Social Games

Social networks have quickly become a standard part of everyday life, and from this social games have gained popularity as well. While not all players of social games play them on Facebook, many do. In a recent study, market research group Lightspeed Research found that 53 percent of the people who use Facebook have played a social game [6]. A study by PopCap Games found that 55 percent of the players are female, with an average age of 48 years old. Less than half, 43 percent, were found to have a college degree, however, an additional 36 percent had completed some college or trade school courses. In the United States, 33 percent of respondents do not play other types of games [7]. Social games are introducing a new breed of players to games who have never played before, as well as, introducing a generation that did not grow up playing games to them.

2 Popular Types of Social Games

In 2009, blogger Tadhg Kelly defined 8 types of Social Games. These included: Social RPGs, Sports RPGs, Casual Games, Word Games, Virtual Villages, Casino, Just for Laughs, and Ownership Games [8]. This set of definitions is problematic, however, in that its reliance on gameplay mechanics to define social games are not critical characteristics of these games as most of them can easily be replicated in a nonsocial system. This definition does not focus on which characteristics or elements make social games unique and how they could be used for learning. To break down social games into discernable categories, it might be more productive to investigate which features propagate the game while leveraging the social aspects of the game.

Social Games gauge success by Monthly Active Unique (MAU) hits, their players who are active in a given month. These players promote to friends by inviting them to play, advertise by posting results to their walls, and otherwise exploit the games they play virally throughout the social network. It is important to not define these games solely by how they promote the game on the social network as most games use multiple strategies to accomplish this point, but how they create returning players by relying on the social aspects. The following represents a list of currently popular or new types of social games that to place reliance on the social network to sustain returning players.

2.1 Energy Depletion Games

Energy depletion games are one of the more popular genres of social games. This type of game is defined by providing the player with a specific amount of energy and allowing them to take actions that use up a predetermined amount of it. Once the energy is depleted to zero, the player must wait for more energy to build up or they must request and/or be gifted energy from their social network before taking more actions [10]. The most popular of these is Zynga's Mafia Wars. In February of 2011, it had 16 million MAUs, more than the entire LinkedIn network [9]. Some activities only require a small amount of energy but larger ones might require the player to wait for hours or even days before taking the action and advancing in the game. As players continue with the game they gain power and build energy faster, making hard tasks easy and impossible tasks attainable.

2.2 Appointment Games

In Appointment Games, the player can immediately start a task in a predetermined location in the game and then wait for their appointment to be rewarded for the task. FarmVille is the most well-known appointment game. Using FarmVille as an example, the player starts with a farm and a limited number of plots in which they can plant seeds. Once the player plants the seeds, they must wait until the plants ripen. Depending upon the particular seeds this could take a few minutes or several days. After the plants have grown, the player must keep their real world time based appointment in order to harvest the plants. If their appointment is missed the plants will wither away and unable to be harvested. Members of an individual's social network may, if so included assist the player by keeping their appointments for them. These types of games have shown to be the most popular on Facebook. The current leader as of February 2011 was CityVille with over 98.6 million MAUs [9].

2.3 Social Competition Games

Social Competition Games follow standard casual game genres in their gameplay style. The most popular game of this type is Bejeweled Blitz, a fast paced high scoring version of the popular casual puzzle game Bejeweled by PopCap Games. These styles might vary greatly between casual puzzle, casino, platformer, or other similar games where points are earned and a competition against their social network friends is created. Players are drawn into this style of game because of the popular draw shown on their network feed. They stay engaged by attempting to beat a previous score or an opposing friend. Further, scores are often reset on a weekly or monthly basis depending upon the game allowing every player within a social network an opportunity to start anew.

2.4 Linked to Friends

The final engagement discussed here is linked to friends. The word "game" is not used, because these engaging activities are often more like toys than actual games. A popular Facebook application of this style is actually titled Zombies. Users might participate in activities related to friends in their network. For example, they might fictionally buy and sell friends to each other. One popular linked to friends application has the user turning friends in to zombies or vampires with no real goal or ambition. Users come back to asynchronously interact with friends or see what their friends have done.

3 Current Educational Examples

Currently, few examples of educational social games that leverage the social network as a tool in a way that could improve learning exist. Social games are being used in education, but these games are primarily social competition or simple casual games that post updates to the user's wall. The following examples detail several applications of social games being used to help students learn:

3.1 Hidden Agenda

Hidden Agenda is a competition for student game developers to create fun games with a hidden educational purpose. The contest has run for many years, but in 2010 the competition was moved to Facebook [11]. With Zynga as a new sponsor, developers were challenged to build educational games for Facebook that will teach high school level material. The first year on Facebook resulted in five games, Dig Deep, AmeriQuiz!, Spider Trig, Timeline the Game, and Body Defense. Each of these games follows the casual games model, and other than posting on a players wall, does very little to tempt the player back.

3.2 EnerCities

EnerCities has incredibly high production value for a social game. Funded by Intelligent Energy Europe, EnerCities has players building a sustainable city. It uses the Unity3D plug-in and contains a high quality 3D world that looks on par if not superior to most commercial entertainment social games. While it appears at first glance to be an appointment game it is actually a social competition game that does not save player progress [12].

This seems to be a missed opportunity to allow the player to see the results of their actions over time. If a player could set up a coal burning plant or a solar energy plant and return after a few days and see the level of pollution and the long-term effect their actions had on the environment they might have learned more than just the short term effect on their current game.

3.3 Farmville for Math Education

An article on eHow.com provides six steps for using Farmville to teach math. The steps are sound, but require paying attention outside of the actual game [13]. Lisa Russell, writing for the home schooling section of Suite101.com suggests there is a simple formula students should learn to determine the Return on Investment (ROI) of actions taken in the game. She states, "Though scientifically inaccurate, Facebook's FarmVille video game offers kids an opportunity to practice math skills, calculating farm profits in a neighborly environment." [14]

Using existing game structures to form lesson plans is not uncommon among games. Students could explore the math in Farmville, but they could certainly do the same thing in other games. The true power of social games for learning is building the learning explicitly into the game itself.

4 Untapped Opportunities for Learning

The current developments in educational social games have not met the potential of the technology. By leveraging the existing paradigms of linked to friends, energy depletion, appointment, and social competition games; educational social games can reach more learners and provide previously unobtainable learning opportunities within social networks. The characteristics of these games that make them appropriate for use in education include, but are not limited to their ability to motivate learners; their ability to make their players feel connected to one another and responsible for their own and in some cases each other's learning and success; and their ability to provide learners with the opportunity to reflect.

4.1 Motivation

One of the most common features of social games is the ability to inform your social network of your achievements and actions in that game. On Facebook, this occurs by posting events to the player's wall. These transmissions of player success are the most important feature for developers to grow their potential customer base by motivating new players to join, and earn similar recognition. In learning games on social networks, this same feature should be leveraged to both allow the player to get recognition for their learning, as well as motivate others to engage in the material.

For example, if a learning game existed on a corporate social network then players could not only share their in-game achievements with their coworkers, but also their boss. The boss would have the ability to see learners progress within the learning game and be informed on their progress. In the right climate, a boss might make a comment on a player's progress and initiative allowing other learners to see that coworkers receiving praise and thereby motivating them to play in order to get similar recognition. Akin to peer pressure, this extrinsic motivator towards learning can be leveraged to bring learners into the content faster and provide a compelling experience at the same time.

4.2 Connected Learning

Players in most social games are provided with opportunities to see each other's progress in the game. This might take the form of a high score table in a social competition game, or actually visiting another person's world in an appointment game. By seeing the progress of friends in their social network, learners have the opportunity to remain connected to the other learners and should be better able to pinpoint their levels of ability and strengths and deficiencies through comparative analysis. The use of casual games in education would also provide opportunities for individual learners to see the climate of their social networks skills and abilities allowing them to reach out to those in need, or when in need to create mentoring types of relationships. The ability to have a transparent learning management system like a social game would also enable students to create their own goals and strategies for their own learning.

4.3 Inter-reliant Learning

Social games are primarily played in an asynchronous multiplayer environment and therefore do not evoke strong similarities to teamwork based learning opportunities. Players and potentially learners in social games are not present in the same place at the same time, but they do have the ability to impact and influence the outcomes of another player's game. Social games prey on their player's sense of responsibility and desire to be a productive member of their social networks by creating situations that a player's chances of success are highly increased if they have the support and assistance of other players. Players can often avoid long waits or periods of inability to play if the other players in their network area ble and willing to complete tasks for them, to give them gifts and to share their own resources.

In Ravenwood Fair players often receive tasks that require a player recruit their social network in order to accomplish, such as having to to restock the assets in your neighbors fair. Players with more neighbors are given more assets daily and are more likely to receive assistance when they request it. In Farmville, if a player misses their appointment another player can fertilize their crops. This provides the player with more time to harvest their crops, it also helps build bonds between players, solidify teams and foster relationships.

4.4 Reflection

In the Army Learning Concept for 2015, the US Army presents a plan to revolutionize learning within the service. One of the issues identified in that document is, "mandatory subjects overcrowd programs of instruction and leave little time for reflection or repetition needed to master the fundamentals . . . training requirements will decrease to avoid increasing the course length while shortening the training day, to allow time for reflection that is necessary to maximize the learning effect." [15] Time for reflection is important for metacognition. J. H. Flavell defines metacognition as, "one's knowledge concerning one's own cognitive processes or anything related to them." [16] Put simply metacognition is a learners self awareness of their own knowledge. Metacognition leads to learner autonomy, independence, and self-regulation.

The appointment game based model encourages players or potential learners to leave the game and return at a later point. During this time the player is aware that the effects of their actions are still ongoing within the social gaming environment.

In FarmVille, in order to grow corn a player simply clicks the plot of land where they want that corn to grow and then must wait for their corn. If more complex task such as selecting the appropriate members for a tactical military unit sending the force on a mission had stop and wait periods built into it; the player could spend their waiting time outside the game reflecting on the decisions they made inside the game.

5 Conclusions

Social games are a new and interesting type of game that have untapped learning potential. Social games for learning will likely never compete with entertainment games in their ability to garner the audience they need to become viral, but they do have characteristics that have positive learning potential. By allowing their users to leverage a social network as part of their learning experience, social games enable multiple types of positive learner attributes, such as self efficacy and self regulation to be elicited. The implementation of social games within learning and social networks contextually tailored towards a specific group, like a company or class might motivate learners to participate in social games. Their ability to connect users as groups or through dependent relationships and the possible mechanic of built in reflection time that can be designed into and/or implemented within social games leaves their potential uncovered. New content and instructional design models that support learning within a social game need to be created and employed into the framework of social games in order to allow accurate research to be executed.

References

1. Rao, V.: Facebook Applications and Playful Mood: the Construction of Facebook as a Third Place. In: MindTrek Conference (2008)
2. Oldenberg, R.: The Great Good Place. Marlow & Company, New York (1999)
3. Facebook Press Room, http://facebook.com/press/info.php?statistics
4. Saleem, M.: Visualizing 6 Years of Facebook, Mashable.com (2010), http://mashable.com/2010/02/10/facebook-growth-infographic
5. TechCrunch, LinkedIn ChrunchBase Profile (2011), http://www.crunchbase.com/company/linkedin
6. Caoili, E.: Study: 53% of Facebook's Users Have Played Social Games, 19% Addicted. Gamasutra: The Art & Business of Making Games (2010)
7. Information Solutions Group, 2010 Social Gaming Research, PopCap Games (2010)
8. Kelly, T.: The 8 Types of Social Game, Gamasutra Expert Blogs Tadhg Kelly (2009)
9. Inside Network, Latest AppData Update: February 2011, AppData: Independent, Accurate Application Metrics and Trends from Inside Network (2011), http://www.appdata.com
10. Sumner, J.: Ideas that will keep players around and coming back. Mafialife Commission (2010)
11. Hidden Agenda Contest., http://www.hiddenagenda.com
12. EnerCities Case Study, Paladin Studios (2011), http://www.paladinstudios.com/cms/case-studies/enercities

Sample Size Estimation for Statistical Comparative Test of Training by Using Augmented Reality via Theoretical Formula and OCC Graphs: Aeronautical Case of a Component Assemblage

Fernando Suárez-Warden, Yocelin Cervantes-Gloria,
and Eduardo González-Mendívil

Instituto Tecnológico y de Estudios Superiores de Monterrey -
Monterrey campus (Monterrey Tech)
Centro de innovación en diseño y tecnología (CIDYT)
Aerospace Competence Development Center (ACDC)
Cátedra de investigación en Máquinas Inteligentes
Av. Eugenio Garza Sada 2501, Monterrey, NL. 64849. Mexico
fersuarezw@gmail.com, yocelincervantes@gmail.com, egm@itesm.mx

Abstract. Advances in Augmented Reality applied to learning the assembly operation in terms of productivity, must certainly be evaluated. We propose a congruent sequence of statistical procedures that will lead to determine the estimated work sample size ($\hat{n}$) according to the level of significance required by the aeronautical sector or a justified similar one and the estimated (sometimes preconceived) value of the plus-minus error (E or E±). We used the Kolmogorov-Smirnov test to verify that a normal distribution fits, it is a nonparametric one (free-distribution). And by taking into account normal population, confidence interval is determined using the Student's t distribution for (n-1) degrees of freedom. We have gotten E error and obtained various sample sizes via statistical formula. Additionally, we proceeded to utilize both an alpha α significance level and a beta β power of the test selected for the aeronautical segment to estimate the size of the sample via application of Operating Characteristic Curves (CCO), being this one of the ways with statistical high rigor. Several scenarios with different $\hat{n}$ values make up the outcome, herein. We disclosed diverse options for the different manners of estimation.

Keywords: Augmented Reality (AR), plus-minus error or margin of error, confidence interval (CI), Operating Characteristic Curves (OCC).

1 Introduction

This document exposes how to determine the work sample size; it consists of a sequence of steps required to find a satisfactory estimated sample size value.

Aeronautical maintenance training is a strategic area within which technicians must go carefully and where people have to take advantage of intelligent technologies such as AR.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 80–89, 2011.

Advances in AR, directed to training for the assembly operations in terms of productivity, need to be evaluated. Since this viewpoint of performance, we think about time invested in training and its return when we use AR to compare with the required standard rate.

Having this in mind, we propose a consistent sequence of statistical steps to get our target of estimating the size of the sample $\hat{n}$ for the level of significance (α) used by the aeronautical sector or a justified similar one and the estimated (properly preconceived) value of the plus-minus error or margin of error (E).

At first we start with the aim of searching such an estimate for this sample size, so one via that we use is a nonparametric method (free-distribution) named Kolmogorov-Smirnov Test, which is performed to verify that a normal distribution fits.

We execute this critical test because it is possible to utilize some known distributions for testing only under normal behavior case, as Statistical Theory states.

After that, taking into account normal population, the confidence interval (CI) is determined by using the Student's **t** distribution for (n-1) degrees of freedom. Work sample sizes for initialization are considered to attain the combined standard deviation $\hat{\mathbf{s}}$ which is used in Student´s **t** distribution to find out several values of the E error.

Consequently we have obtained the E error to get the estimated sample size $\hat{n}$ via statistical formula utilized in the theory developed for these cases. We can hold the above paragraph when we have a small sample size case. And this is the situation here.

We are going to generate several scenarios with the same level of significance or probabilities of accomplish type I error (α) and confident values of E for the previous $\hat{n}$ calculations in order to make up the searched outcome.

Then we should be careful in estimating the sample size to be representative and avoid fall not only in lack but also in excess.

1.1 We Propose to Use via Theoretical Formula and OCC Graphs

Being rigorist, we build a methodology for providing two different ways:

a) Formula derived from the confidence interval assuming normal population
b) Graphs of Operating Characteristic Curves (OCC)

According to Kazmier [4], for normal population we use theoretical formula for sample size (n) based on the population variance and margin of error, but first we must do an adaptation to the estimation of n and the estimated value of sampling standard deviation **S**.

Besides, we must meet the requirements for alpha α significance level (or probability of type I error) and beta β power of the test (or probability of type II error) selected for the aeronautical segment to state the estimated sample size via graphics of Operating Characteristic Curves (CCO), being this, one of the techniques of high statistical rigor.

2 Paper Preparation

Now we acquire the data achieved from Mercado [6]: the assemblage times by using RA for a wing of an aircraft RV-10 kit. Below is a table with the results obtained in his thesis.

Table 1. Results from the AR method

EXPERIMENT	X = TIME (min)	ERRORS (qty)	QUESTIONS (qty)
1	248	3	4
2	210	2	3
3	236	2	3
4	**284**	2	**5**
5	197	0	1
6	166	0	2

Source: Mercado[6]. *Improving Training and Maintenance. Operations in Aeronautic Related Processes Using Augmented Reality.*

2.1 Data Analysis and Filtering

With the aim of avoid slanted sampling or skewed data generation that would lead to a bad estimation of S, we will accomplish the next data analysis:

Mercado [6] shows in pp. 32 his *Table 3* where only the variables *Time* and *Questions* are centered on *Boxplot diagram* (from Minitab) while the variable *Errors* is biased.

Also, in his *Table 1*, he only take into account the columns *Time* and *Questions* that appear as Graphics 1st. and 3rd. in *Table 2* (pp. 30 and 31 respectively) and here is where we note with interest the last point that corresponds to his: EXPERIMENT #4 with *Time 284, Questions 5.*

Filtering. Because that is the farthest point from the straight line of the statistical trial of Kolmogorov-Smirnov Normality test and therefore we will remove it and determine well, in a filtered form, a more representative average than that obtained by Mercado [6]. In a second step we will eliminate and substitute such a point.

2.2 Formulas

Permit us present several formulas:

1. We obtain the average of the data Xav with the following formula

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

2. The sample variance

$$s^2 = \frac{\sum_{i=1}^{n}\left\{(x_i - \bar{x})^2\right\}}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2}{n-1} \qquad (2)$$

3. And based on discrete uniform distribution, from Eq. (3.7) in Appendix:

$$u^2(x_i) = \frac{1}{12}(a_+ - a_-)^2 \qquad (3)$$

where: α_+ = MAX , $\alpha\_$ = MIN, and $\hat{s}^2 = u^2$

4. The sampling standard deviation

$$s = \sqrt{s^2} \qquad (4)$$

5. The Combined Std. Deviation

$$\hat{\mathbf{s}}_D = \hat{\mathbf{s}} / n^{1/2} \qquad (5)$$

6. Weighed Combined Std. Deviation

$$\hat{\mathbf{s}}_{Xav1 - Xav2} = \hat{\mathbf{s}}_1 \, (\mathbf{1/n_1 + 1/n_2})^{1/2} \qquad (6)$$

7. In order to obtain the estimation of **d**:

$$đ = (1/2\, \hat{\mathbf{s}}_D) \, | Xav1 - Xav2 | \qquad (7)$$

8. From OCC graph we get $\hat{n} = (n^* + 1) / 2$ (8)

   for a given value of β (= 0.01 in aeronautical sector).

Other formulas shall appear according to paper development. For the aviation sector, we take the level of significance α = 1% = 0.01, thus: α/2= .005

From statistical tables for normal distribution, we find $z_{\alpha/2}$ $Z_{.005} = 2.58$

To begin with, we initially make calculations, by mean of previous formulas, using filtered Mercado´s data (see **Filtering** at section **2.1**), which are shown below:

Reminder: we removed the farthest point from the straight line of the statistical tests of Kolmogorov-Smirnov Normality test and therefore the average changes to:

Xav =Average ($\bar{x}$)	211.4
Variance (s^2)	1053.8
Std. Deviation (s)	32.462
Combined Std. Deviation ($\hat{\mathbf{s}}_D$)	= 27/ (5)½ = 12
d estimated = đ	= [(½)/ (12)] l 243 – 211.4l = 1.32

Where we considered X = TIME.

2.3 Kolmogorov- Smirnov Test

We will validate if the experimental data model fits the normal distribution statistical model, by executing the Fit goodness Kolmogorov- Smirnov Test:

Ho: The spent time follows a normal distribution

Hi: The spent time does not follow a normal distribution

To determine should the obtained and filtered data follow a normal distribution and are significant accomplish the Kolmogorov- Smirnov (K-S) test. This K-S test indicates whether and how well the model fits the data.

The K-S statistical test calculates the distance between the theoretical normal distribution and the distribution estimated from the residuals, adjusted for suspensions, and reports the p- value (specific probability value). When the p-value is less than α (1%) the null hypothesis is rejected, and accepted otherwise.

Once the K-S test was applied to several selected cases to be discarded someone or various, we achieved an acceptable fit goodness for: Normal distribution and Gamma distribution.

In our case, the results were a p-value > α (= 1%) for the K-S fit goodness test:

Ho: normal distribution fits
H1: normal distribution does not fit

And therefore we could not reject Ho, so we got to assume normality. Please, see Hogg and Tanis [3] who developed examples of K-S test. Note: Gamma distribution fits too.

2.4 Confidence Interval

According to Kazmier [4] we can apply t-student test in order to determine the confidence interval (CI), by using the next expression for experimental values (texp):

$$texp = (\bar{x}_1 - \bar{x}_2) / \hat{s} (1/n_1 + 1/n_2)^{½} \tag{9}$$

And we can build the confidence interval CI for small random sample using the t distribution with degrees of freedom ν = n1 + n2 - 2 (assuming and / or having tested normal population by mean of fit goodness via K-S test). From:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\nu,\alpha} \hat{s} (1/n_1 + 1/n_2)^{1/2} \tag{10}$$

we get:

$$E = CI / 2 = t_{\nu,\alpha} \hat{s} (1/n_1 + 1/n_2)^{1/2} \tag{11}$$

3 Via Theoretical Formula and OCC Graphs

After it has been proved our population or universe were normal it should be used with all forcefulness the corresponding theoretical formula of Kazmier [4]:

$$n = (z\,\sigma / E)^2 \tag{12}$$

which we adapt for dealing with an estimated value:

$$\hat{n} = (z\,\hat{s} / E)^2 \tag{13}$$

where $\hat{s}$ is the estimator for σ, E is the margin of error and $\hat{n}$ is the estimator for n.

3.1 Scenarios Developed by Formula

It was stated that: $\hat{n} = (z\,\hat{s} / E)^2 = \hat{s}^2 (z / E)^2$ from Eq. (13) We knew the result $\hat{s} = 32.462$ based on Std. Deviation (s) formula.

The Confidence Interval CI is depicted for small random sample (assuming and / or having tested normal population fit goodness, via K-S Test).

1. Reminder 1: we removed the farthest point from the straight line of the K-S normality test (see **Filtering** at section **2.1**) to get a 1st new margin of error E. Such as Eq. (10):

$$\bar{x} \pm t_{v,\alpha}\ \hat{s}/n^{1/2} \tag{14}$$

$\bar{x} \pm t_{4,\,0.01}\ \hat{s}/n^{1/2} = 211 \pm (4.604)\ 32.462\ /5^{1/2} = 211 \pm 66.83;\ E = 66.83$

By introducing attained data into adapted formula:

$\hat{n} = (z\,\hat{s} / E)^2 = \hat{s}^2\ (z / E)^2 = 1053.8(2.58 /66.83)^2 = 1.57 \approx 2$; $\hat{n}$ is an integer.

2. Reminder 2: we removed the farthest point from the straight line of the K-S normality test (see **Filtering**) and substituted it to get a modified average and a 2nd representative margin of error E:

$\bar{x} \pm t_{5,\,0.01}\ \hat{s}/n^{1/2} = 211 \pm (4.032)\ 23.67\ /6^{1/2} = 211 \pm 38.95;\ E = 38.95$

By introducing attained data into adapted formula:

$\hat{n} = (z\,\hat{s} / E)^2 = 560.3\ (2.58 /38.95)^2 = 2.45 \approx 3$

3. For two depending samples (associated pairs), where $\nu = n_1 + n_2 - 2 = 5 + 5 - 2 = 8$

$\bar{x} \pm t_{8,\,0.01}\ \hat{\mathbf{s}}_1\ (1/ n_1 + 1/ n_2)^{1/2} = 211 \pm (3.5)\ 14.97 = 211 \pm 52.4;\ E = 52.4$

By bringing together obtained into adapted formula:

$\hat{n} = (z\,\hat{s} / E)^2 = 560.3\ (2.58 /52.4)^2 = 1.35 \approx 2$

3.2 Scenarios Developed by OCC Charts

4. With the resultant quantities gotten by formulas (2.2 section), we reached the first amount for đ (= 1.32). And placing this estimated value of **d** as input into CCO chart:

For $\beta = .01$ from OCC graph we get $\tilde{n}^* = 15$

Then, $\hat{n} = (\tilde{n}^* + 1) / 2 = (15 + 1) / 2 = 8$ by Eq. (8)

5. After that, we detached the farthest point from the straight line of the K-S normality test (see **Filtering**) and therefore other representative average resulted:

Xav modified = 211.4 and $\hat{\mathbf{s}}_2$ modified = 32.462

$\hat{\mathbf{s}}_1 = 23.67$ based on discrete uniform distribution (rectangular distribution)

And we assumed $\hat{\mathbf{s}}_1 \approx \hat{\mathbf{s}}_2 = \hat{\mathbf{s}} = 23.67$

On the other hand, Combined Std. Deviation $\hat{s}_D = \hat{s} / n^{½} = 23.67/ (5)^{½} = 10.6$
So: $đ = [1/ 2(10.6)] |243 - 211.4| = 1.4925 \approx 1.5$
With this value as input in CCO chart, for a beta β = .01, from OCC graph:
We get $ñ^* = 14$ and $\hat{n} = (ñ^* + 1) / 2 = (14 + 1) / 2 = 7.5 \approx 8$ by Eq. (8)

6. We removed the farthest point from the straight line of the K-S normality test (see **Filtering**) and substitute it for the previous average to get another new Xav (= 211.3)

And another new Combined Std. Deviation $\hat{s}_D = \hat{s} / n^{½} = 23.67/ (6)^{½} = 9.66$
So, we have a new: $đ = [1/ 2 (9.66)] |243 - 211.3| = 1.6384$
With this value as input in CCO chart, for a beta: β = .01 from OCC graph:
We get $ñ^* = 8$ and $\hat{n} = (ñ^* + 1) / 2 = (8 + 1) / 2 = 4.5 \approx 5$ by Eq. (8)

7. Also, for depending samples (associated pairs), a weighed Combined Std. Deviation:

$$\hat{s}_{Xav1 - Xav2} = \hat{s}_1 (1/n_1 + 1/n_2)^{½} = 14.97 \qquad \text{from Eq. (6)}$$

Thus we exploit it in determining $đ = 1/ 2 (14.97)] |243 - 211.3| = 1.058$
For β =.01, from OCC graph $ñ^* = 21$, so $\hat{n} = (ñ^* + 1) / 2 = (21 + 1) / 2 = 11$ by Eq.(8)

4 Results

The scenarios we obtained for n̂ were:

Table 2. Various scenarios of estimated sample size (for depending and independent sample types)

Scenario	**Sample type**	**Via**	**ŝ**	**z**	**E error**	**Estimation n̂**	**α**	**β**
1.-	independent	Formula	32.462	2.58	66.83	2	.01	-
2.-	independent	Formula	23.67	2.58	38.95	3	.01	-
3.-	depending	Formula	14.97	2.58	52.4	2	.01	-
Scenario	**Sample type**	**Via**	**$\hat{s}_D$**		**d**	**Estimation n̂**	**α**	**β**
4.-	independent	CCO	12	-	1.32	8	.01	.01
5.-	independent	CCO	10.6	-	1.4925	8	.01	.01
6.-	independent	CCO	9.66	-	1.6384	5	.01	.01
7.-	depending	CCO	14.97	-	1.058	11	.01	.01

By the next manage approvals, we briefly want to clarify the OCC charts use. Input:

a) Set α = 0.01 (aeronautical sector), hence select the correspondent OCC chart
b) Enter to horizontal axis by addressing the estimated value for **d:**

$$đ = (1/ 2\hat{s}_D) |Xav1 - Xav2| \qquad \text{from Eq. (7)}$$

c) Select the β = 0.01 (aeronautical sector) on the vertical axis of the OCC chart.

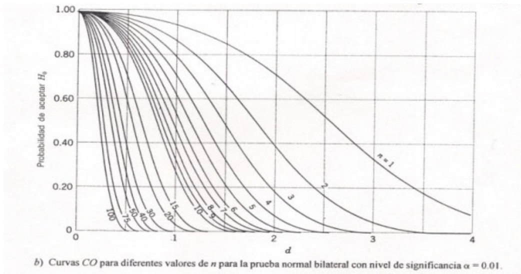


Fig. 1. Operating Characteristic Curves (OCC). Source: Hines and Montgomery [2]

Output:

- ñ* value in OCC chart
- with ñ* thus we estimate the work sample size $\hat{n} = (\tilde{n}^* + 1) / 2$ by Eq. (8)

5 Conclusion

We learn to solve assessment sample size problems by doing research in advance so necessary, especially when the situation demands to know a good approximation of the corresponding magnitude before running the experiment. This is our study case, herein.

Several scenarios of estimated sample size could be construed from two points of view.

We think about different manners for getting the two outcome types. However it is crucial to consider the beta probability (β) to achieve our target given that we also must try to avoid falling in the type II error. For this reason, the second way gave more credible results for the proposed sizes at 4, 5, 6 and 7 rows of Table 2 by utilizing the CCO via.

We invite you to take advantage of this study which is focused on estimating the size for small work samples in order to cut costs within an adequate criterion.

Acknowledgments. We want to thanks to professors Salvador García-Lumbreras, Ph.D., José Guadalupe Ríos, Ph.D. and María del Carmen Temblador, Ph.D. from Monterrey Tech for their support and recommendations.

References

1. Campbell, J.D., Jardine, A.K.S.: Maintenance Excellence: Optimizing Life-Cycle Decisions. Marcel Dekker, USA (2001)
2. Hines, W.W., Montgomery, D.C.: Probabilidad y Estadística para ingeniería y administración. In: CECSA (2005)
3. Hogg, R.V., Tanis, E.A.: Probability and Statistical Inference. Macmillan, NYC (1977)
4. Kazmier, L.: Estadística Aplicada a la Administración y la Economía. McGrawHill, NewYork (1993)

5. Mendenhall, W., Schaffer.: Estadística Matemática con Aplicaciones. Grupo Editorial Iberoamérica (1990)
6. Mercado Field, E.R.: Improving Training and Maintenancce Operations in Aeronautic Related Processes Using Augmented Reality. ITESM- Monterrey. México (2010)
7. Organización Internacional para la Estandarización (ISO). Estándar Int. ISO 3534-1 Estadísticas –Vocabulario y símbolos-Parte I: Probabilidad y Estadística en gral., 1a ed. Génova, Suiza (1993)

Appendix: Discrete Uniform Model (Rectangular Distribution)

Organización Internacional para la Estandarización (ISO) [7].

Disclaimer. This section is composed by the international standard ISO 3534-1 Vocabulary and Symbols-Statistics, Part I. It is an official translation from ISO organization, so sorry we are not authorized for converting to English. Let us use Spanish.

Distribución uniforme discreta (distribución rectangular). (*)El uso correcto de la información disponible para la evaluación Tipo B de la incertidumbre estándar en la medición tiene un adentramiento basado en la experiencia y conocimiento general. Es una habilidad que puede ser aprendida con la práctica. Una evaluación de Tipo B de la incertidumbre bien basada puede ser tan confiable como la de Tipo A, especialmente en una situación de medición donde la valoración Tipo A es fundada sólo en un pequeño número comparativo de observaciones estadísticamente independientes. Casos a ver:

a) Sólo un *valor singular* es conocido por la cantidad X_i, p. ej. un valor sencillo medido, un valor resultante de una medición previa, un valor de referencia de literatura o un valor de corrección; usado para x_i. La incertidumbre estándar $u(x_i)$ asociada con x_i debe ser adoptada cuando es dada. De otro modo, es calculada de datos de incertidumbre inequívocos, pero si no están disponibles, ésta debe evaluarse basándose en la experiencia.

b) Cuando la *distribución de probabilidad* puede ser asumida de la cantidad X_i, basada en teoría o experiencia, el valor esperado y la raíz cuadrada de la varianza deben ser tomados como un estimado de x_i y la incertidumbre estándar asociada $u(x_i)$ respectivamente.

c) Si sólo los *límites inferior y superior* a_+ y a_- pueden ser estimados para el valor de la cantidad X_i (por ejemplo, las especificaciones de la manufacturera de un instrumento de medición, un rango de temperatura, un error de redondeo o de truncado resultado de una reducción automática), una distribución de probabilidad con una densidad de probabilidad constante entre esos límites (distribución de probabilidad rectangular) tiene que ser asumida para la posible variabilidad de la cantidad de entrada X_i. Según el caso (**b**):

$$x_i = \frac{1}{2}\left(a_+ + a_-\right)$$ para el valor estimado Xav o Xprom; y: (3.6)

$$u^2\left(x_i\right) = \frac{1}{12}\left(a_+ - a_-\right)^2 \quad (3.7)$$

para el cuadrado de la incertidumbre estándar $\hat{\mathbf{s}}^2$. Donde: $a+$ = MAX , $a-$ = MIN.

Enhancing English Learning Website Content and User Interface Functions Using Integrated Quality Assessment

Dylan Sung

Department of Applied Linguistics and Language Studies
Chung Yuan Christian University, 200 Chung Pei Rd., Chung-Li, Taiwan 32023
dsung@cycu.edu.tw

Abstract. The present study investigated the applicability of an integrated quality assessment approach to assess English learning website quality. The study used the Kano Model to identify attractive quality attributes of the content and user interface functions of an English learning website. The Importance-Satisfaction Model was used to determine the interface functions that need to be improved. Findings of the study led to the conclusion that the content and user interface functions of English learning websites should be specially developed according to the satisfaction level of the learners and also the degree of importance perceived by them. On the basis of the key quality attributes identified by utilizing the integrated quality assessment model developed in this study, English learning website designers can make important decisions on specific areas for enhancing the quality of the website and improving the learning efficiency of the users.

Keywords: English as a foreign language (EFL), English learning, computer-assisted language learning (CALL), Internet-assisted language learning, e-learning, educational technology.

1 Introduction

With the advent of the digital age and the proliferation of educational technology, effective integration of e-learning can be very advantageous in higher education [1]. As the current learning websites are being updated at a torrid pace, educators and website designers should continuously assess the quality of the latest technology from different aspects [2]. In particular, technical improvements in the actual learning content and interface functions can give rise to enhanced learning outcomes.

In the CALL context, many e-learning websites have been developed worldwide, but little research has been conducted concerning the development of comprehensive evaluation criteria [3]. To develop students' necessary English skills in the global information society with the most convenient and practical means, it is of significance to further improve the existing Internet-assisted language learning tools. These tools already possess the practical advantage of being readily available to the students. In spite of the huge demand for user-friendly language learning websites and materials, most of them still lack standardized management and consistent quality at the current time [4]. The vast amount of learning materials in e-learning has also brought about

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 90–99, 2011.

the challenge of using suitable learning materials for a particular learning topic, creating the need for user recommendation within a learning context [5]. What English learning websites require is constant feedback from the perspective of the users in context, so that users' quality of learning can be improved [6].

In the broad field of quality assessment, many methods are available for investigating requirement characteristics of the users. Of the range of methods available, one of the widely utilized methods by academic researchers and industries is the Kano Model. A few examples of domains that the practical research tool has been applied to include website design [7], web-based learning [8], and product design optimization [9]. In the EFL context, Sung [10] conducted a cross-national study that utilized the Kano Model to identify attractive quality attributes of English language teaching at two universities from two East Asian countries.

Whilst the Kano Model is an established tool for sorting quality attributes into various categories, the comparatively new Importance-Satisfaction Model (I-S Model) can be considered a simple yet powerful tool for finding out the *excellent* attributes and the *to be improved* attributes [11]. Analytical results derived from the I-S Model allow identification of quality attributes that need to be improved upon. This key function provides important guidelines for quality decisions. Based on the analyses and decisions, required actions can then be initiated for quality improvement.

Two of the most important dimensions of an e-learning website are its content quality [12] and user interface design [13]. Therefore, English learning website designers committed to providing an effective language learning experience to users should not only focus on the aesthetics and the sheer amount of learning material that they can place on a website. Instead, they should also highlight the strengths and weaknesses of the existing design and the content area for improvement. In view of these critical aspects, integrating the Kano Model and the I-S Model with the aim of obtaining more valuable quality information becomes a strategic move.

Taking into account the functional versatility of the Kano Model and the useful applications of the I-S model, it can be presumed that the integrated approach holds promise as a fitting tool for assessing website quality in the virtual English learning environment. For these reasons, the Kano Model and the I-S Model were chosen and integrated to be applied to this study.

The purpose of this study was to investigate the applicability of an integrated quality assessment approach to assess English learning website quality via user feedback information. With the aim of achieving the stated purpose, the following objectives were formulated to guide the study:

1. To examine students' perceptions regarding English learning through web-based e-learning;
2. To classify attractive learning content;
3. To determine interface functions that require immediate improvement.

2 Methodology

This section presents the research design and procedures for the study. First, the participants' background information will be introduced. Next, the learning content of the English learning website and instrumentation used in this study will be illustrated. Lastly, the procedures of data collection and data analysis will be presented.

2.1 Population and Sample

The target population for this study represented all university freshmen from a selected university in northern Taiwan during the Spring semester of 2010. Cluster sampling was employed to randomly select six classes from six different colleges with varying majors. This method is appropriate for educational research situations where the population members are grouped in classes. The sample participants for the investigation included 219 freshmen.

2.2 Weekly Use of the Website

All participating students used the online English website under study weekly for 18 weeks throughout the Spring 2010 semester. The learning content of the website is divided into five thematic units including daily life, living environment, transportation, recreation, and holidays and festivals. The units contain anywhere from seven to 13 lessons for a total of 52 different topics altogether. The intention behind this design was for students to learn one lesson per week throughout the year. For this study, students were randomly assigned two lessons per week for a total of 36 lessons.

Students were asked to use all eight interface functions under investigation including vocabulary, pictures, pronunciation, related phrases, dialogue practice, slang, idioms, and test to learn the materials. Moreover, in-class quizzes specifically designed to reinforce consistent use of the eight interface functions were administered on a weekly basis.

2.3 Instrumentation

A survey instrument was designed to collect data for this study. The items of the survey developed were in accordance with the actual learning content and interface functions of the website. The survey contains 39 items and is divided into four parts: (1) English learning websites, (2) quality attributes of website content, (3) quality attributes of the interface functions, and (4) importance level and satisfaction level of the interface functions. There are 18 items in the first part, five items in the second part, eight items in the third part, and eight items in the fourth part.

In each item on the instrument, there is a five-point Likert-type scale for participants to choose from. Part 1 of the instrument uses a scale which ranges from 5 (strongly agree) to 1 (strongly disagree). Part 2 and part 3 were structured in reference to the Kano Model. These parts use a five-point Likert-type scale which ranges from 1 (very satisfied) to 5 (very unsatisfied). Part 2 aims to investigate how students would feel if a particular thematic unit was either implemented or unimplemented. Part 3 aims to investigate how students would feel if a particular interface function was either implemented or unimplemented. Each item in part 2 and part 3 requires the participants to provide a pair of responses. The first in each pair, a functional item, refers to a situation in which the thematic unit or interface function is implemented.

The second in the pair, a dysfunctional item, refers to an opposite situation in which the same attribute is unimplemented.

Part 4 of the questionnaire is structured in reference to the I-S Model. Similar to parts 2 and 3, this part also uses a five-point Likert-type scale (1, very important to 5, very unimportant for the degree of importance and 1, very satisfied to 5, very unsatisfied for the satisfaction level). The questionnaire aims to identify quality attributes of interface functions that need immediate improvement. Each item in this part requires the participants to provide a pair of responses as well. The first in each pair deals with the degree of importance of the quality attribute. The second deals with the satisfaction level of the quality attribute.

2.4 Data Collection

At the end of the 18-week semester, the researcher distributed the survey questionnaires to the 219 students from the six classes under study. The researcher specified that the results derived from the questionnaires were going to be reported for research purposes only and no individuals were going to be identified. The time needed to complete the questionnaires was approximately 15 minutes. Proper ways to complete the questionnaire, especially parts 2, 3, and 4, were carefully explained by the researcher.

With the intention of obtaining the best results, the students were given sufficient time to complete the questionnaires. The researcher also encouraged the students to carefully select their best response. All questionnaires were returned directly to the researcher for data analysis.

2.5 Data Analysis

Upon return of the survey questionnaires, all data were checked first to detect if they were usable. Of the 219 questionnaires returned, Two (0.91%) were found to be incomplete and four (1.83%) contained invalid data that could not be categorized. Two-hundred and thirteen (97.3%) were complete and useable for data analysis. All usable data from the questionnaires were then entered and analyzed.

For objective 1, descriptive analyses, including means and standard deviations, were performed to examine students' perceptions regarding English learning through web-based e-learning. For Objective 2, thematic units and interface functions were analyzed using the Kano Model [14]. Using English learning websites as the setting, thematic units or interface functions can be categorized based on the interplay of the student's responses for the thematic unit or interface function when present and when absent into the five following categories:

- Attractive quality attribute: A thematic unit or interface function that gives satisfaction if present, but that produces no dissatisfaction if absent.
- One-dimensional quality attribute: A thematic unit or interface function that is positively and linearly related to student satisfaction – that is, the greater the degree of implementation of the thematic unit or interface function, the greater the level of student satisfaction.

- Must-be quality attribute: A thematic unit or interface function whose absence will result in student dissatisfaction, but whose presence does not significantly contribute to student satisfaction.
- Indifferent quality attribute: A thematic unit or interface function whose presence or absence does not cause any satisfaction or dissatisfaction to students.
- Reverse quality attribute: A thematic unit or interface function whose presence causes student dissatisfaction, and whose absence results in student satisfaction.

Table 1. Kano Evaluation Table

Attribute Category / Implementation	Attractive	One-dimensional	Must-be	Indifferent	Reverse
Implemented	1	1	2, 3, 4	2, 3, 4	5
Unimplemented	2, 3, 4	5	5	2, 3, 4	1

The thematic units and interface functions were categorized in accordance with the Kano Evaluation Table (Table 1). A quality attribute falls into two categories if the difference between the highest rated attribute and the second highest rated attribute is less than 3%.

For Objective 3, interface functions were analyzed using the I-S Model. Interface functions in need of immediate improvement were determined using the I-S Model. In the I-S Model, the horizontal dimension shows the degree of importance of a particular interface function, and the vertical dimension shows the satisfaction level of the interface function [11]. The order pair (importance scale and satisfaction scale) can then be located on the coordinates. The means of the importance scale and the satisfaction scale can be then used to divide the coordinate into the four following areas: excellent, to be improved, surplus, and care-free.

3 Findings

The results of data analysis of the study are presented in this section. The main strengths and weaknesses of the website under study are presented first. This section also reveals the various quality attributes of thematic units and interface functions derived from the data analyses. In the Kano Model, thematic units or interface functions that fell into the *attractive* category are analyzed first, followed by ones that fell into the *one-dimensional* and *must-be* categories. *Indifferent* quality attributes are reported last. In the I-S Model, interface functions that fell into the *to be improved* area are analyzed first, followed by ones that fell into the *excellent*, *surplus*, and *care-free* categories.

3.1 English Learning Website

The following results were derived from descriptive statistics. Students agreed with the flexibility in time of use of the learning website (M = 4.42, SD = 0.62). The ability for repeated practice on learned material provided by the website (*M* = 4.31, *SD* = 0.63) was indicated as the second advantage. The convenient easy access to use the website (*M* = 4.26, *SD* = 0.67) was indicated as another positive trait.

From a learning standpoint, the overall content of the website was considered practical for English learning (*M* = 3.76, *SD* = 0.69) to a certain extent. From a technical standpoint, the general interface design of the website was considered easy to use (*M* = 3.92, *SD* = 0.64).

The lowest rated item was that the website's design can lead to improvement in speaking skills (*M* = 2.43, *SD* = 1.04). Likewise, the students did not agree that the website's design can lead to improvement in writing skills (*M* = 2.49, *SD* = 0.98). In contrast, the students agreed that the website's design can lead to improvement in listening comprehension (*M* = 3.51, *SD* = 0.99) and reading comprehension (*M* = 3.54, *SD* = 0.97).

3.2 Quality Attributes of Website Content in Kano Model

Categorized data of the website content quality attributes identified by the students are reported in Table 2. Ninety-five (44.60%) students identified Daily Life as an *attractive* quality attribute whereas 89 (41.78%) students identified Recreation as an *attractive* quality attribute. The other three thematic units were indentified as *indifferent* quality attributes.

Table 2. Categorization of Website Content Quality Attributes

Unit/ Attribute	Attractive		One-dimensional		Must-be		Indifferent		Reverse		Attribute Category
Daily Life	**95**	**44.60%**	25	11.74%	31	14.55%	62	29.11%	0	0.00%	**Attractive**
Living Environment	62	29.11%	22	10.33%	27	12.68%	102	47.89%	0	0.00%	**Indifferent**
Transportation	43	20.19%	45	21.13%	19	8.92%	106	49.77%	0	0.00%	**Indifferent**
Recreation	**89**	**41.78%**	49	23.00%	20	9.39%	55	25.82%	0	0.00%	**Attractive**
Holidays and Festivals	54	25.35%	27	12.68%	21	9.86%	111	52.11%	0	0.00%	**Indifferent**

3.3 Quality Attributes of Interface Functions

Categorized data of the website interface function attributes identified by the students are reported in Table 3. Of the eight interface functions, Related Phrases was found to be the most *attractive* quality attribute by 102 (47.89%) students. Pronunciation was identified as an *attractive* quality attribute by 75 (35.51%) students. Additionally, Dialogue Practice was identified as *attractive* by 84 (39.44%) students and as *indifferent* by 78 (36.62%) students.

There was only one interface function, Vocabulary, that 73 (34.27%) the students identified as a *one-dimensional* quality attribute. The remaining four interface functions were indentified as *indifferent* quality attributes.

Table 3. Categorization of Interface Function Attributes in Kano Model

Function/ Attribute	Attractive		One-dimensional		Must-be		Indifferent		Reverse		Attribute Category
Vocabulary	43	20.19%	**73**	**34.27%**	46	21.60%	51	23.94%	0	0.00%	**One-dimensional**
Pictures	48	22.54%	41	19.25%	17	7.98%	107	50.23%	0	0.00%	**Indifferent**
Pronunciation	**75**	**35.21%**	35	16.43%	44	20.66%	59	27.70%	0	0.00%	**Attractive**
Related Phrases	**102**	**47.89%**	28	13.15%	21	9.86%	62	29.11%	0	0.00%	**Attractive**
Dialogue Practice	**84**	**39.44%**	29	13.62%	22	10.33%	**78**	**36.62%**	0	0.00%	**Attractive/ Indifferent**
Slang	51	23.94%	30	14.08%	18	8.45%	114	53.52%	0	0.00%	**Indifferent**
Idioms	36	16.90%	22	10.33%	29	13.62%	125	58.69%	1	0.47%	**Indifferent**
Test	25	11.74%	16	7.51%	39	18.31%	131	61.50%	2	0.94%	**Indifferent**

3.4 Importance Level and Satisfaction Level of the Interface Functions

Categorization results derived from the I-S Model are presented in Figure 1. The mean degree of importance for the eight interface functions was 3.99. The mean satisfaction level was 3.45. The mean degree of importance is used as the horizontal axis. In addition, the mean satisfaction level is used as the vertical axis. The two axes divide the coordinate into four quadrants.

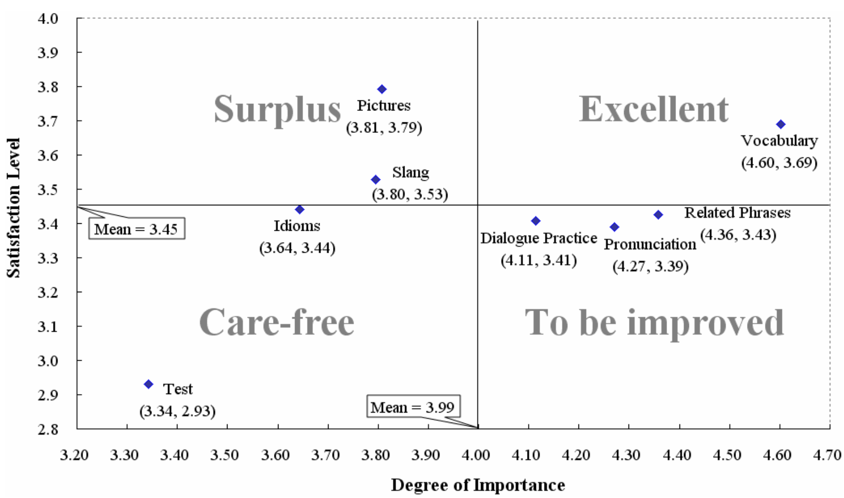


Fig. 1. Categorization of Interface Function Attributes in the I-S Model

The results indicate that there are three interface functions located in the *to be improved* area. Based on the ranking of importance, they are related phrases, pronunciation, and dialogue practice. Vocabulary is the only function in the *excellent* area. Pictures and slang are in the *surplus* area. The remaining two functions, idioms and test, are in the *care-free* area.

4 Discussion

Students have positive perceptions toward the flexibility, repeated practice, and convenient easy access offered by the English learning website. The overall content of the English learning website was found to be of practical use. Students perceived daily life as the most *attractive* thematic area to learn using the website. Since the students have to deal with academic and technical English in the classroom setting, they seem to be more inclined to acquire everyday English in their own virtual learning environment. The general interface design was deemed satisfactory. Still, its limited usefulness for users to practice output skills of speaking and writing was pointed out as a major weakness. Designers of English learning websites should focus on this technical limitation and improve it in order to meet current user needs.

Dialogue Practice received polarized responses. While it was identified as the second most *attractive* user interface function, it was also identified as *indifferent* by approximately one-third of the students. What's more, it fell into the *to be improved* area with relatively low ratings. Likewise, Related Phrases and Pronunciation were also identified as *attractive* functions that fell into the *to be improved* area. These three desirable yet underperforming functions can affect users' development of oral communication competency. The current limitation of the technology to effectively help students practice speaking skills calls for the need to improve the interactive content of the website. Within the past decade, several large-scale language proficiency tests worldwide started to implement mandatory speaking test tasks to comprehensively assess four skills [15, 16]. In increasingly more non-English speaking countries, certain proficiency tests have been imposed on millions of university students as one of their degree requirements [17, 18]. Considering the aforesaid issues and the importance of adequate speaking skills to produce accurate communication, there is an apparent need to tackle the technology's current inability to functionally develop speaking skills. After all, as important as vocabulary is in the acquisition of a foreign language, speaking is as crucial to effective communication.

The Test function was found to be the most *indifferent*, unimportant, and unsatisfied function by a great majority of the students. This finding reflects Kartal's [19] claim that most language learning websites do not utilize the Internet to its full potential, and the pedagogical scenarios and learning theories are reduced to just answering structural exercises such as multiple choice questions, true or false items, and fill in the blanks. Moreover, the proficiency levels, the learning objectives, and the target user group are not clearly specified in these websites. The Test function of the website under study contains only 10 vocabulary matching questions per lesson. The overly simplistic structure may be seen as a cause of the negative user feedback. For online tests to reflect the range of language skills the website intends to offer, they should be designed to measure more than vocabulary acquisition. However, it has been advocated that Internet-based language tests should employ simplified interfaces if students are to be tested in their knowledge of the target language as opposed to being tested on their performance and skills as computer users [20]. To strike a desirable balance, a variety of tests should be developed to assess different language skills while retaining user-friendliness.

Due to the nature of the Kano Model and I-S Model, all data of this study were limited to Likert-type responses. Subsequent studies should include relevant open-ended questions to gather additional data. Interviews should also be conducted to elicit information regarding the preferred improvement that users would like to see.

5 Conclusions

The present study successfully utilized the Kano Model and the I-S Model to assess quality attributes of an online English learning website. Findings of the study led to the conclusion that the content and user interface features of English learning websites should be specially developed according to the satisfaction level of the learners and also the degree of importance perceived by them. On the basis of the attractive attributes identified by the students, instructors can make appropriate changes to the lessons they assign to the students. This will likely increase student motivation as they would be learning what they consider attractive. Degree of importance and satisfaction level of the users can also update website designers with significant information regarding particular interface functions that require immediate improvements. It is recommended that the integrated approach (Figure 2.) developed in this study can be used to enhance the quality of the content and specific user interface functions of English learning websites.

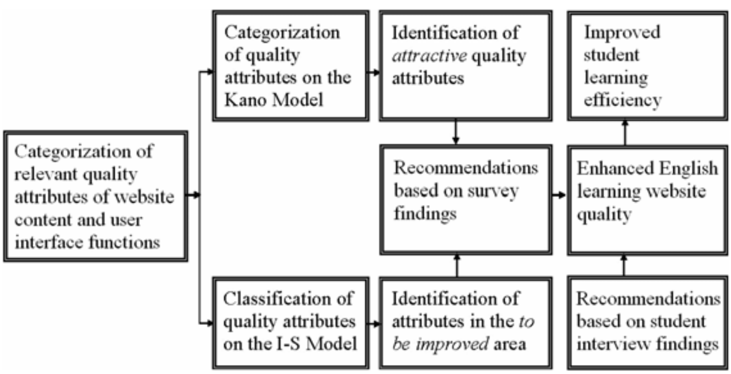


Fig. 2. The Model of Enhancing English Learning Website Quality Using Integrated Quality Assessment

References

1. Ellis, R., Goodyear, P.: Students' Experiences of e-Learning in Higher Education: The Ecology of Sustainable Innovation. Routledge, New York (2009)
2. McNaught, C., Lam, P.: Building an Evaluation Culture and Evidence Base for E-Learning in Three Hong Kong Universities. British Journal of Educational Technology 36(4), 599–614 (2005)

3. Liu, G.Z., Liu, Z.H., Hwang, G.J.: Developing Multi-Dimensional Evaluation Criteria for English Learning Websites with University Students and Professors. Computers & Education 56(1), 65–79 (2011)
4. Kartal, E., Uzun, L.: The Internet, Language Learning, and International Dialogue: Constructing Online Foreign Language Learning Websites. Turkish Online Journal of Distance Education 11(2), 90–107 (2010)
5. Ghauth, K.I., Abdullah, N.A.: Learning Materials Recommendation Using Good Learners' Ratings and Content-Based Filtering. Educational Technology Research and Development 56(8), 711–727 (2010)
6. Reeves, T.C., Hedberg, J.G.: Interactive Learning Systems Evaluation. Educational Technology Publications, New Jersey (2003)
7. von Dran, G.M., Zhang, P.Z., Small, R.: Quality Websites: An Application of the Kano Model to Website Design. In: Proceedings of the Fifth Americas Conference on Information Systems, Milwaukee, pp. 898–900 (1999)
8. Lee, W.I., Shih, B.Y., Tu, L.J.: The Application of Kano's Model for Improving Web-Based Learning Performance. In: 32nd Annual Frontiers in Education, 1, T3E2–32, Boston (2002)
9. Lai, X., Xie, M., Tan, K.C.: Optimizing Product Design through the Kano Model and QFD. In: Proceedings of the IEEE International Engineering Management Conference, Singapore, vol. 3, pp. 1085–1089 (2004)
10. Sung, D.: Attractive Quality Attributes of English Language Teaching at Two East Asian Universities. The Journal of Asia TEFL 6(4), 131–149 (2009)
11. Yang, C.C.: Improvement Actions Based on the Customers Satisfaction Survey. Total Quality Management and Business Excellence 14(8), 919–930 (2003)
12. Lee, I.: Korean Content Management in e-Higher Education: Here and Hereafter. Educational Technology Research and Development 54(2), 209–219 (2009)
13. Nam, C.S., Smith-Jackson, T.L.: Web-Based Learning Environment: A Theory-Based Design Process for Development and Evaluation. Journal of Information Technology Education 6, 23–43 (2007)
14. Kano, N., Nobuhira, S., Takahashi, F., Tsuji, S.: Attractive Quality and Must-Be Quality. Journal of Japanese Society for Quality Control 14(2), 39–48 (1984) (in Japanese)
15. Stoynoff, S.: Recent Developments in Language Assessment and the Case of Four Large-Scale Tests of ESOL Ability. Language Teaching 42(1), 1–40 (2009)
16. Powers, D.E.: The Case for a Comprehensive, Four-Skills Assessment of English-Language Proficiency. R&D Connections. Educational Testing Service, vol. 14. Princeton (2010)
17. Shih, C.M.: The Washback of the General English Proficiency Test on University Policies: A Taiwan Case Study. Language Assessment Quarterly 7(3), 234–254 (2010)
18. Zhang, Y., Elder, C.: Measuring the Speaking Proficiency of Advanced EFL Learners in China: The CET-SET Solution. Language Assessment Quarterly 6(4), 298–314 (2009)
19. Kartal, E.: The Internet and Autonomous Language Learning: A Typology of Suggested Aids. The Turkish Online Journal of Educational Technology 4(4), 54–58 (2005)
20. Garcia Laborda, J., Magal-Royo, T., de Sigueira Rocha, J.M., Alvarez, M.F.: Ergonomics Factors in English as a Foreign Language Testing: The Case of PLEVALEX. Computers & Education 54(2), 384–391 (2010)

The Influence of Virtual World Interactions toward Driving Real World Behaviors

Hari Thiruvengada, Paul Derby, Wendy Foslien, John Beane, and Anand Tharanathan

Honeywell (ACS) Advanced Technology Labs
1985 Douglas Drive N., Golden Valley, MN 55442, USA
{Hari.Thiruvenagda,Paul.Derby,Wendy.Foslien,John.Beane,
Anand.Tharanathan}@Honeywell.com

Abstract. In the recent years, virtual worlds have gained wide spread popularity and acceptance in a variety of application domains including training, education, social networking, and conceptual demonstrations. This is largely due to their ability to support modeling of fully textured high-resolution real world objects, to provide a compelling user experience, and to offer novel, rich and exploratory interactions to the user. However, the impact of familiarity with the real world domain and objects on user behavior is still unclear. In this study, we discuss the findings from a pilot study on a virtual world facility tour that was based on a real world facility. The objectives of the tour were threefold. First, we sought to understand the feasibility of using a virtual tour in lieu of the actual real world tour. Second, the tour was used as an educational tool to demonstrate several sustainable or efficient energy initiatives to the facility occupants. Specifically, the virtual tour consisted of an interactive energy dashboard, a low voltage LED based lighting demonstration, an illustration of Heating, Ventilation and Air Conditioning (HVAC) equipment operations during day and night, and renewable energy sources within the facility. Third, we sought to understand the impact of the tour on participants' future behaviors and attitudes toward sustainable energy. In order to address these overarching objectives, user feedback was collected using a survey after the users participated in the tour. We administered the survey to both occupants and non-occupants of the facility to also understand the impact of familiarity on their behaviors. Users who were familiar with the facility were optimistic about their perception of learning how to navigate around the virtual replica than those who were not familiar. Our preliminary findings from the survey indicate that virtual worlds can have a positive impact on the users' behavior. Overall, we found that users' engagement during the virtual tour could contribute to learning and the development of lasting positive behaviors within virtual world, which can, in turn, translate into real world behaviors.

Keywords: Virtual Worlds, Experiential Learning, Human-in-the-loop simulation.

1 Introduction

Virtual worlds are one of the fastest growing technology thrust areas in gaming and social networking communities. They have been applied rigorously toward training

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 100–109, 2011.

and experiential learning applications in the recent years (Cohn, Nicholson & Schmorrow, 2009; Nicholson, Schmorrow & Cohn, 2009; Schmorrow, Cohn, & Nicholson, 2009; Tharanathan, Derby & Thiruvengada, in press; Thiruvengada, Tharanathan & Derby, in press). With the availability of large network bandwidths, high end computing resources, advanced voice and video communications, user experience has become more lively, engaging and quite seamless within online communities. This has led to the development of many virtual worlds with varying levels of fidelity. In this research, we attempt to understand the impact of a user's virtual world interactions and experience on real world behaviors within the context of an energy sustainability initiative. We propose that a compelling interaction bundled with an engaging user experience can have a lasting impact on user's perception and encourage energy conservation behaviors.

2 Literature Review

A virtual environment (VE) is a simulated, computer generated environment that creates "synthetic sensory experiences" (Salas, Oser, Cannon-Bowers and Daskarolis-Kring, 2002). For example, users could interact with computer-generated images, sounds, and haptics. More specifically, multiuser virtual environments (MUVEs), also referred to as virtual worlds (VW), allow users to interact with other users within the simulated environment for the purpose of collaboration and knowledge sharing (Bainbridge, 2007; Salas, Oser, Cannon-Bowers & Daskarolis-Kring, 2002). Oftentimes, virtual words mimic complex physical environments and are inhabited by other users represented by animated characters, or avatars (Bainbridge, 2007). Two examples of VWs are Second Life® [1] and OLIVE™ [2].

Virtual worlds have proven to be successful within a variety of domains such as corporate training and collaboration (Heiphetz & Woodill, 2009), military training (Shines, 2002), medical training (Heiphetz & Woodill, 2009), prototype development (Linden Labs, 2009), psychological treatment (Anderson, Rothbaum & Hodges, 2001), and higher education (Wankel & Kinsley, 2009). This success can be attributed to the virtual world's ability to provide the user an immersive environment that is suitable for role-playing, problem-solving, collaboration, experiential learning, and interaction (Heiphetz & Woodill, 2009). That is, virtual worlds allow users to develop cognitive skills and acquire knowledge while being situated within an environment that mimics a real, inaccessible, and/or dangerous environment (Dieterle & Clarke, 2008; Mantovoni, 2001).

3 Virtual World Demonstration Framework

Our framework was developed within a virtual world called Second Life® to educate the visitor about the sustainability and energy initiatives that were being implemented at our facility in real world. We called it the "Golden Valley Virtual Tour" (GVVT). We adopted an iterative user centered design approach to construct the components of

[1] Second Life® is a registered trademark of "Linden Research, Inc."
[2] OLIVE™ is a trademark of "SAIC, Inc." (Formerly owned by Forterra Systems, Inc.)

our GVVT framework. The details of the user centered design approach and the individual tour stop components of our GVVT framework are discussed below.

3.1 User Centered Design

In order to engage the user effectively and deliver a compelling user experience, we followed a user centered design approach as described by Usability Professionals' Association (UPA) guidelines and User Centered Design Methodology (Mayhew, 1999). We adhered to the key steps identified within the four distinct phases: *Analysis*, *Design*, *Implementation* and *Deployment*. During the *Analysis phase*, we met with the key stakeholders including building managers, champions/members of the sustainability committee, public relations to identify the vision for this demonstration. We also identified key usability goals and objectives and transformed them into needs and requirements. A field survey of the real world facility was conducted to capture the details so that we could use them in the design and implementation phases. At this point, we identified several user profiles but narrowed it down to three key user profiles (*facility occupant, building manager, executive member*) and created detailed personas. A task analysis was conducted to identify user scenarios and user performance requirements. In the *Design Phase*, we began brainstorming design concepts and metaphors and came up with four key tour stops (*Energy Dashboard; Motion Detected LED Lighting; HVAC Day/Night Modes; Wind Turbine and Solar Panel*) for the demonstration. The specific details of each tour stop are discussed further in the rest of this section. The demonstration included an automated and interactive robot tour guide who was modeled after a real life tour guide to walk the visitor through the various tour stops in a pre-determined sequence. We developed detailed workflow and navigation for the users based on the use case scenarios identified earlier. Then, we proceeded to develop storyboards, interaction paradigms, and robot tour guide scripts for each tour stop. We documented the scripts using a standard template that was specifically designed for this demonstration. During the *implementation phase*, we conducted ongoing evaluation of the implemented prototypes during recurring meetings with the development team within the virtual world facility. We also adopted an iterative build approach, wherein, each tour stop was developed first and perfected before we constructed the next tour stop. An informal heuristic evaluation and usability testing of each tour stop was also conducted. This enabled us to work closely with the development team as the design was implemented and improve the quality of the tour stops significantly. In the final *Deployment Phase*, we conducted reviews with the stakeholders and received feedback from them using surveys to understand the effectiveness of the tour stop to achieve the goals and objectives of the user when they engaged with this demonstration framework.

3.2 Description of Framework Components

The GVVT framework components consists of the virtual facility that is a recreation of the real world facility, a robotic tour guide and four key tour stops (*Energy Dashboard; Motion Detected LED Lighting; HVAC Day/Night Modes; Wind Turbine* *and Solar Panel*) that are organized in a hierarchical manner. The hierarchical organization of our GVVT framework is shown in Fig. 1.

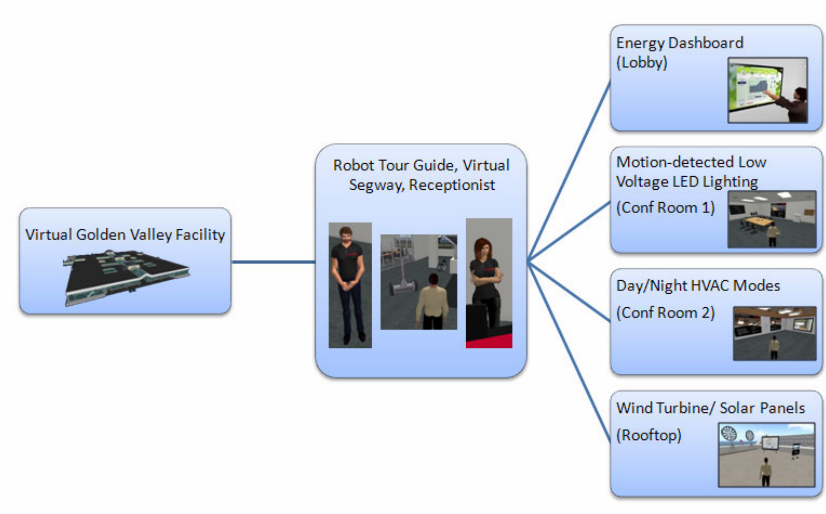


Fig. 1. Hierarchical Organization of GVVT Framework

A virtual representation of Golden Valley facility was developed. The robot tour guide that resembles a human avatar was responsible for directing the visitor through the various tour stops in a predetermined sequence. The role of the robot tour guide can be considered similar to the role a tour guide at a museum. To visit the individual tour stop, the visitor would have to first register with the robot tour guide and instruct it to proceed to the subsequent tour stop in order to experience the same. In addition, we built an alternative fully automated tour experience using a Virtual "Segway® Personal Transporter[3]", where the visitor would not have to stop and interact with the robot tour guide to proceed to the next tour stop. We also constructed an interactive and automated receptionist avatar capable of greeting the visitors (based on proximity) and providing key facts about the virtual facility when the visitor first logs into the virtual tour. This framework is robust, scalable and can be easily customized to include additional tour stops in the future. The tour stops were classified into four groups based on the mode of user interaction and level of feedback as shown in Fig. 2.

The visitors were able to interact manually with the objects of interest within tour stops 1 & 3 using their avatars either via keyboard or mouse. In tour stop 2 & 4, the location of the visitor's avatar and the proximity to objects of interest triggered interactions (e.g. Motion-detected Low Voltage LED lighting). Similarly, the visitors received limited (direct) feedback in tour stops 1 and 2 through audio narration and states of individual objects. In tour stops 3 & 4, the visitors received more (indirect) feedback through illustrative videos, in addition to direct feedback about object states and through audio narration.

---

[3] Segway® Personal Transporter is a registered trademark of "Segway, Inc."

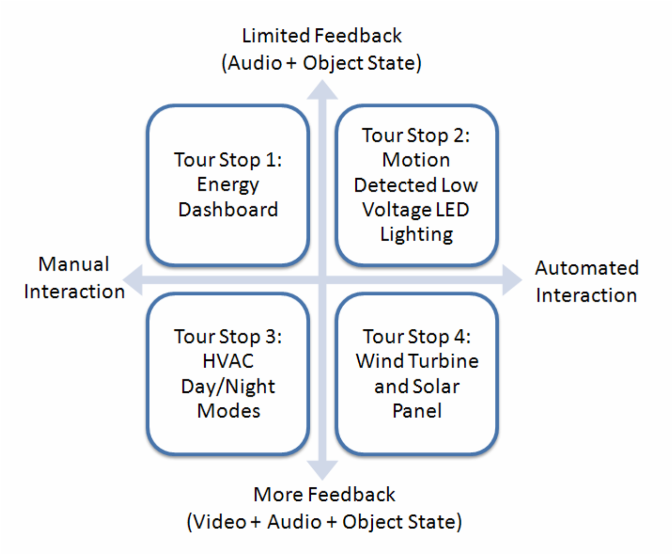


Fig. 2. Classification of Tour Stops

3.3 Tour Stop 1: Energy Dashboard

The focus of this tour stop was to inform the visitor about the current energy consumption of the building and educate them about the energy conservation measures. We composed a storyboard and script that would be narrated by the robot tour guide to inform the visitor about the significance of this tour stop in addition to user interaction methods. To achieve this goal, we used an interactive web based dashboard and superimposed that on a virtual large screen display object within Second Life®. We enabled the user's avatar to interact with the dashboard contents and review information such as current energy usage, historical energy consumption and comparison for each section of the facility, etc. This tour stop aimed at raising awareness on how the building is performance from an energy consumption perspective and how users can contribute to global energy conservation both within and outside the facility. A snapshot of the energy dashboard is shown in Fig. 3.

3.4 Tour Stop 2: Motion-Detected Low Voltage LED Lighting

In this tour stop, we aimed to inform the user about the ongoing initiative to reduce energy consumption using motion-detected low voltage light emitting diode (LED) lighting source. We composed a storyboard and script that would be narrated by the robot tour guide to inform the visitor about the significance of this tour stop in addition to user interaction methods. Each conference room is equipped with motion detected lighting controls. Sensors automatically detect when the room becomes occupied and turns on the lights. When occupants exit the room, the sensors no longer detects motion and turn off after a small period of time. To interact with this tour stop the visitor would control his/her avatar and enter the conference room which would automatically turn the lights on. The layout of this tour stop is shown in Fig. 3.

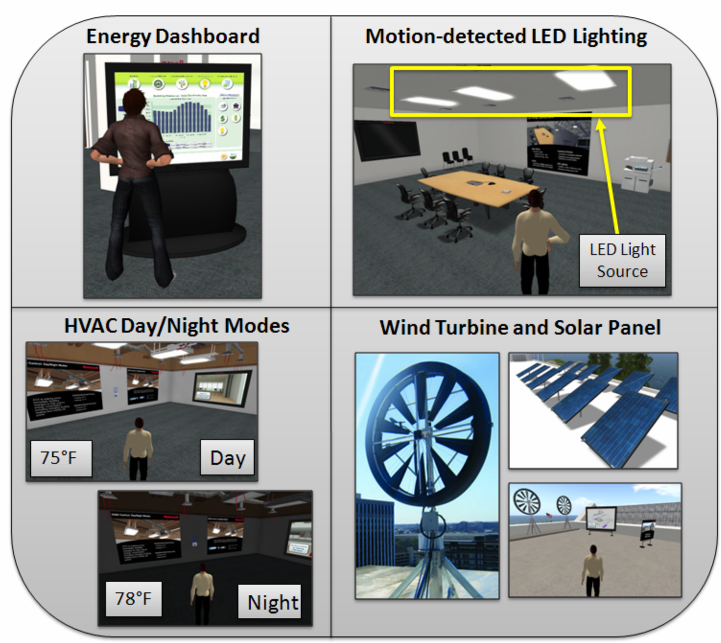


Fig. 3. Snapshot of Tour Stops 1-4: Energy Dashboard (*Top Left*); Motion-detected Low Voltage LED Lighting (*Top Right*); HVAC Day Night Modes (*Bottom Left*); Wind Turbine and Solar Panel (*Bottom Right*)

This change alone has been estimated to save roughly $1000 per hour for all of the conference and training rooms combined. In addition, the lighting in conference rooms, have been replaced with LED lighting, which is estimated to have an approximate energy savings of roughly 43%. In other areas of the building, such as in the atrium, daylight sensors detect when there is enough ambient lighting provided from the skylights. When there is not enough natural ambient lighting, then the atrium lights will turn on. In open office areas, there is also a time based lighting control mechanism that automatically turns the lights on during normal operations and turns the lights off during after hours.

3.5 Tour Stop 3: HVAC Day/Night Modes

The goal of this tour stop is to inform the visitors about the operations of the HVAC equipment and how it impacts the controls during day time (when the facility is occupied) and night time (when the facility is unoccupied). We created a storyboard and script that would be narrated by the robot tour guide to inform the visitor about the significance of this tour stop in addition to user interaction methods. It informs the visitor that the HVAC control system attempts to reduce energy consumption at night when the facility in not in use by lowering the temperature set points during the warmer months and raising the temperature set points during the colder months. By doing this, we are not continuously heating or cooling unoccupied areas of the building. It also informs the user to see how the temperature of the building changes from day to night on the virtual wall mounted digital thermostat in addition to the air flow from the ducts. In addition, a short video clip that illustrates the HVAC operations is played on a large screen display in a looped manner. This enables the user to interact, explore and learn about the HVAC operations. A snapshot of this tour stop is shown in Fig. 3.

3.6 Tour Stop 4: Wind Turbine and Solar Panel

In this tour stop, we intended to highlight the wind turbine and solar panel green energy initiatives and the offsetting benefits for using such green energy generation technology. Currently, we have installed and are testing a WindTronics®[4] wind turbine on the roof of the real world facility and replicated a similar concept within the virtual facility. It is being used to power the lighting within a corridor of our facility. This tour stop also includes an illustrative short video clip that visually informs how the wind turbine operates. We provided examples of our vision for installing solar panel arrays on the roof top and educated users the potential benefits. A snapshot of this tour stop is shown in Fig. 3.

4 Method

Participants: We conducted a preliminary pilot study using two different user groups to understand the usability and perceived impact of GVVT framework tour stops on their long term energy consumption behaviors. Group 1 had 23 participants and was called “GV”. The members of this group were occupants of the real world facility and were very familiar with the physical layout and interior/exterior of the building. They had limited exposure to Second Life® and passively observed the avatar’s interactions at the tour stops. Group 2 was called “DLF” and had 13 participants. The members of this group were primarily instructional designers and had some exposure with Second Life®. They were not occupants of our facility and were shown the building layout only through floor plans, photographic images of the interior and maps of the exterior of the building. They passively observed the avatar’s interactions at the tour stops.

Survey: Both groups were administered a brief survey and were asked to rate the survey results based on a Likert rating scale of 1(Strongly Disagree) to 5 (Strongly Agree). The key results for the survey questions are discussed in the next section.

5 Results

The questions posed to the two groups resulted in some common themes, and also some differences between the two groups. Fig. 4 shows a normalized histogram to illustrate the differences in responses between the GV and DLF group respondents for the questions that were asked of both groups.

---

[4] WindTronics® is a registered trademark of “WindTronics LLC”.

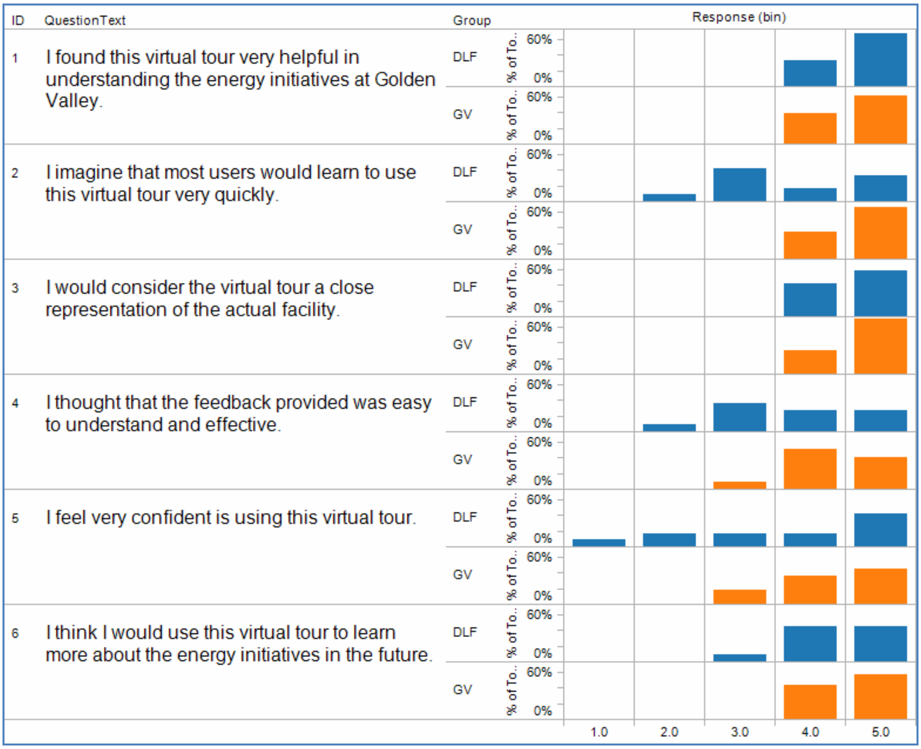


Fig. 4. Normalized Histogram of Survey Responses

Both groups found the tour helpful in understanding the energy initiatives, considered the tour a close representation of the facility, and thought they would use a virtual tour to learn more in the future. One difference noted between groups is in the level of confidence in using the tour. The DLF group tended to have less confidence in using the tour on their own. A second area of difference was in the perception of how users would learn to use the virtual tour, where the DLF group was more skeptical that end users would quickly learn. The third area of difference was in the perception of whether the feedback was easy to understand. Here, the DLF group had more spread than GV group, and also provided more comments and questions on the feedback.

The results suggest that non-familiarity with the real world facility may have negatively impacted the perceived ratings of some subjects for questions 2 (learning), 4 (ease of understanding) & 5 (confidence). Perhaps, this may be attributed to the cognitive dissonance between real world and virtual world infrastructures. Since the DLF group has never seen our facility before except in screenshots during the presentation, they were being extremely conservative in their ratings. On the other hand, GV Occupants felt that they could compensate easily based on their familiarity with the real world facility. The causal factors for this phenomenon need to be further validated with long term studies and more subjects.

Another observation is that the mode for GV group on question 2 (learning) & 4 (ease of understanding) appears to be much higher than the DLF group. A possible cause is the DLF group had an instructional design background and felt that there is a need to improve the learning deployment mechanisms, ease of understanding and boost the confidence of the user in the actual tour.

The mode for question 1 (learning) & 3 (representation) appears to be higher than the average rating (3) for both groups, which indicates that the test subjects were optimistic about being able to use this framework as a learning tool and thought that the virtual world facility closely resembled the real world facility.

Acknowledgements. We would like to thank the occupants of the Honeywell Golden Valley facility and the members of Digital Learning Forum (DLF) for their survey responses. DLF (http://www.digitallearningforum.com/) is Special Interest Group of the Minnesota Chapter of International Society for Performance Improvement (MNISPI). For more information about MNSIPI, go to http://www.mnispi.org.

References

1. Anderson, P.L., Rothbaum, B.O., Hodges, L.: Virtual reality: using the virtual world to improve quality of life in the real world. Bulletin of the Menninger Clinic 65, 78–91 (2001)
2. Bainbridge, W.S.: The scientific research potential of virtual worlds. Science 317, 472–476 (2007)
3. Cannon-Bowers, J.A., Bowers, C.A., Sanchez, A.: Using synthetic learning Environments to train teams. In: Sessa, V.I., London, M. (eds.) Work Group Learning: Understanding, Improving and Assessing how Groups Learn in Organizations. CRC Press, Boca Raton (2007)
4. Cohn, J., Nicholson, D., Schmorrow, D.: Integrated systems, training evaluations, and future directions. In: The PSI handbook of virtual environments for training and education: developments for military and beyond. Praeger Security International, Westport, CT (2009)
5. Dieterle, E., Clarke, J.: Multi-user virtual environments for teaching and learning. In: Pagani, M. (ed.) Encyclopedia of multimedia technology and networking, 3rd edn. Idea Group, Hershey (2008)
6. Heiphetz, A., Woodill, G.: Training and collaboration with virtual worlds: How to create cost-saving, efficient, and engaging programs. McGraw Hill, New York (2009)
7. Lampton, D.R., Bliss, J.P., Morris, C.S.: Human performance measurement in virtual environments. In: Stanney, K.M. (ed.) Handbook of virtual environment: Design, Implementation, and Applications. Erlbaum, Mahwah (2002)
8. Linden Labs. Simulation training and prototyping in virtual worlds: Northrup Grumman in Second Life. (retrieved from 2009), `http://secondlifegrid.net/casestudies`
9. Mantovani, F.: VR learning: potential and challenges for the use of 3D environments in education and training. In: Riva, G., Galimberti, C. (eds.) Towards cyberpsychology. IOS Press, Amsterdam (2001)
10. Mayhew, D.J.: The usability engineering lifecycle: A practitioner's handbook for user interface design (interactive technologies). Morgan Kaufmann Publishers, San Francisco (1999)

11. Nicholson, D., Schmorrow, D., Cohn, J.: The PSI handbook of virtual environments for training and education: Developments for military and beyond. VE Components and Training Technologies, vol. 2. Praeger Security International, Westport (2009)
12. Riva, G.: Virtual reality in neuro-psycho-physiology: cognitive, clinical and methodological issues in assessment and rehabilitation. IOS Press, Amsterdam (1997)
13. Salas, E., Oser, R.L., Cannon-Bowers, J.A., Daskarolis-Kring, E.: Team training in virtual environments: An event-based approach. In: Quiñones, M.A., Ehrensstein, A. (eds.) Training for a Rapidly Changing Workplace: Applications of Psychological Research. APA, Washington (2002)
14. Schmorrow, D., Cohn, J., Nicholson, D.: The PSI handbook of virtual environments for training and education: Developments for military and beyond. Learning, requirements, and metrics, vol. 1. Praeger Security International, Westport, CT (2009)
15. Tharanathan, A., Derby, P., Thiruvengada, H. . Training for metacognition in simulated environments. In: Rothrock, L., Narayan, S. (eds.) Human-in-the-loop simulation: Methods to practice. (in press)
16. Thiruvengada, H., Tharanathan, A., Derby, P.: PerFECT: An automated framework for training on the fly. In: Rothrock, L., Narayan, S. (eds.) Human-in-the-loop simulation: Methods to practice,(In Press)
17. Wankel, C., Kinsley, J.: Higher Education in Virtual Worlds: Teaching and learning in Second Life. Emerald Group Publishing Limited, UK (2009)

Interactive Performance: Dramatic Improvisation in a Mixed Reality Environment for Learning

Jeff Wirth[1], Anne E. Norris[2], Dan Mapes[3], Kenneth E. Ingraham[4], and J. Michael Moshell[5]

[1] Wirth Creative, Orlando, Florida USA
[2] College of Nursing, University of Central Florida, Orlando, Florida USA
[3] Institute for Simulation & Training, U. of Central Florida, Orlando, FL USA
[4] Interactor Simulation Systems, Inc., Orlando, Florida USA
[5] School of Visual Arts & Design, U. of Central Florida, Orlando, Florida USA
jwirth@wirthcreative.com, anorris@mail.ucf.edu, dmapes@ist.ucf.edu, kenneth.ingraham@gmail.com, jm.moshell@cs.ucf.edu

Abstract. A trained interactive performer uses a combination of head-motion capture and a new desktop gesture/posture control system to enact five avatars on a screen, as those avatars interact face-to-face with a participant/trainee. The inter-actor, assisted by a narrator/operator, provides voices for all five on-screen characters and leads the participant through a story-driven improvisational experience. This paper focuses on the processes of scenario development, inter-actor training and production management in the new creative discipline of interactive performance in mixed reality environments.

Keywords: interactive performance, mixed reality, avatar, role-playing, learning, training.

1 Introduction

1.1 Interactive Performance

Interactive Performance (IP) is an emerging discipline in which trained inter-actors work with an untrained spect-actor ('spect') to improvise a story. Beginning the mid-1990's, Jeff Wirth has developed a body of theory and practice and a training methodology for preparing inter-actors [1]. Five areas of skill and knowledge are required for a successful inter-actor: story structure, psychology, improvisation, dramatic performance and technology. During the interval between 2000 and 2010, Wirth trained a group of inter-actors at the University of Central Florida and conducted a series of projects and experiments to apply their skills to tasks related to entertainment, learning and training. Wirth's StoryBox was featured at the 2009 Chicago Improv Festival.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 110–118, 2011.

In interactive performance, there is no traditional script. The inter-actors must be free to respond as the situation develops. However, the team needs to be able to structure the story experience and plan their actions and reactions. To a first-time spect-actor (participant) in an interactive performance, the sequence of events may seem entirely spontaneous. The 'spect' may say or do things which directly and immediately impact the development of the story. However, the trained inter-actors are continuously applying a rich mix of psychological skills, knowledge of story structure and experience in improvisation to shape and support the spect's experience.

1.2 The Latina Empowerment Project

Anne Norris and Charles Hughes (UCF Computer Science and Institute for Simulation & Training) received funding from the National Institutes of Health to develop a game using avatars and real-life scenarios to build skills in Latina middle-schoolers for resisting peer pressure to engage in sexual behavior. They assembled a team and set forth to create a system that uses one or more inter-actors to play the roles of peers, interacting with the targeted teenage participants. The key idea is to create a simulated peer group interaction within a game-like context. The inter-actors control cartoon-like characters on a computer screen. A story-driven dialog with several scenes leads the participant to experience peer pressure, and provide opportunity to practice peer resistance skills.

This paper focuses on the issues concerned with interactive performance in mixed reality environments. The research objectives, methods and results concerning training to resist peer pressure, and the details of the specific scenario will be published elsewhere. We describe the characters within the scenario because they are particularly germane to the process of interactive performance.

2 Story Development and Training

2.1 The Development Process

Jeff Wirth was engaged to develop the content of the game. The overall strategy was as follows:

- Recruit a team of inter-actors to conduct focus groups and field work, and to use these experiences to develop characters and a story through a series of improvisation sessions.
- Use PowerPoint to produce simple visuals for the characters and situations in the story, and video projection to test the story with a few typical Latina teenagers through a *live storyboarding* process.
- Document the results and provide guidance to the programming and art team, so as to construct an avatar-based system for formal experiments.
- Conduct a series of rehearsals using the deliverable version of the game, with three dimensional avatars.
- Test the final version with students in a middle school.

2.2 Field Work

A team of three women with prior inter-actor or theatrical experience were recruited to accomplish this work. In March and April 2010, two focus groups, each involving approximately ten Latina middle school girls, were held. In May 2010, Wirth involved the inter-actors in field work at the middle school. They observed and interacted with girls and boys in the after school program, observing interpersonal non-verbal and verbal behaviors.

The team then met with two high school students who were Latina for ten design sessions of approximately four hours each, during June and July. They used improvisational techniques to develop middle school characters and a story. This process was facilitated by initially providing archetype characters, which the group then named and described.

The game's characters:

Vicki - a friend and companion to the protagonist, also a sixth grader.
Julissa - the coolest girl in the class; physically the most mature; an antagonist and source of temptation
Margarita - younger, less developed. The 'goat' of the group.
Javier - the coolest boy in the seventh grade and source of temptation
Zack - the slightly younger and goofier sidekick to Javier

The protagonist is the participant. She has no on-screen representation. Rather, she is looking at, and being 'seen' by the on-screen characters. (Actually she is being seen by the inter-actor(s) controlling the characters, via a video link.) This mixture of real and virtual presence has been termed mixed reality [2].

2.3 Live Storyboarding

A PowerPoint presentation was constructed, built around simple flat cartoon-like pictures of the five characters listed above. At this stage, three inter-actors were used, although the ultimate plan was to have one inter-actor control all the avatars. The three inter-actors, a group of observers and three Latina teen-agers conducted a pilot testing process. We describe this process as "live storyboarding", because the PowerPoint scenes serve a role similar to that of storyboards in planning a movie. They allow experimentation with the layout of scenery, positioning of the characters in scenes, and their respective entrances, interactions and exits.

2.4 Physical Setup

A conference room at the Institute for Simulation and Training, approximately 20 feet wide and 40 feet long, has an airwall that divides it into two rooms. One of these rooms was set up with a rear projection screen. Between this screen and the far wall was a table and chair, observed by a video camera on a tripod. The airwall was partially closed. On a table sitting in the gap in the airwall was a projector, which illuminated the back of the rear projection screen. The participant (i.e., middle school girl playing the game) sat at the table facing the screen.

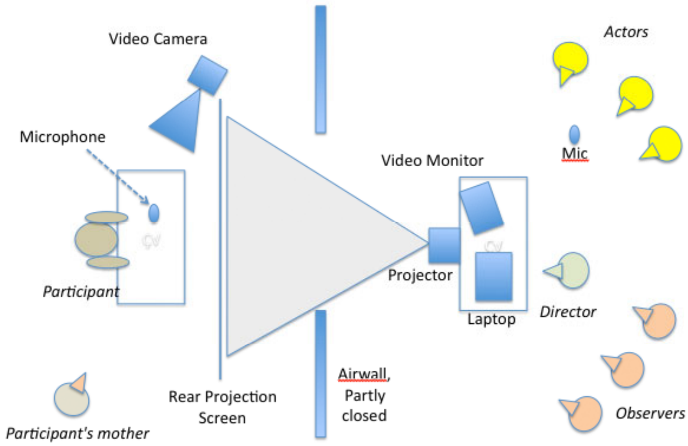


Fig. 1. Setup for the Live Storyboard Tests

In the other room, sitting on a table was a control console which contained a video monitor (displaying an image of the seated participant), and an audio mixer (not shown in diagram). This system was set up so that the show director was sitting at the table, facing through the gap in the airwall, and watching the back side of the rear projection screen. To his right were three actors, gathered around a microphone. They could see the participant's video image and the cartoon characters on the screen of the laptop. They could also see the rear projected imagery, but neither they nor the show director (Wirth) could directly see the participant.

The inter-actors spoke into a shared microphone; the sound was reproduced by speakers in the participant's room, but she could also hear the actors' voices through the open airwall door. Likewise, the participant could hear the actors' voices partially through the airwall. There was no delay or other impediment to the audio 'boost' system, as there is no Skype or other digitization of the audio being done.

2.5 Trial Runs

Two trial runs with participants were staged. The participants and their parents were interviewed, and the team then planned revisions to the storyline based on the experience. As the personalities of the five characters evolved, the inter-actors gained confidence in their ability to portray them. Parts of the scenario that 'worked' were expanded. The entrances and exits of the characters were modified. In one case, an entire scene was excised and replaced by a different scene, to remove an ambiguity about what happened. In the original plan, a 'fade to black' allowed the participant to imagine what took place between one scene and the next. This kind of openness to spect-actor creativity is common in interactive performance, but it did not serve the didactic purposes of the current project. It created too much ambiguity on the part of the inter-actor as to what the spect imagined was happening.

3 Transition from Live Storyboarding to 3d Avatar System

3.1 A Novel Control Mechanism

A novel control mechanism was designed by Dan Mapes for control of the 3d avatars. Derived from the TeachMe system [3] for remote avatar control, the new system differs in several fundamental ways. TeachMe, like the new system, enables a single inter-actor to control five avatars. However, TeachMe is a totally analog system. The inter-actor wears a motion-capture shirt and hat, and controls five seated avatars (one at a time). The hand, torso and head motions of the inter-actor are directly mapped onto the motion of the avatars. A relatively simple system provides plausible movements to the other avatars while the inter-actor has control of a specific avatar.

The new (patent-pending) system could be described as a "five dimensional" control metaphor, representing a blend of discrete and continuous control. The inter-actor moves a pawn-like control device (with a motion capture ball on its head) across a two dimensional map on the desktop to select an avatar and a particular group of gestures or behaviors. These behaviors are custom-animated to support the scenario being enacted. Figure 2 shows the controls used in the first demonstration/tests.

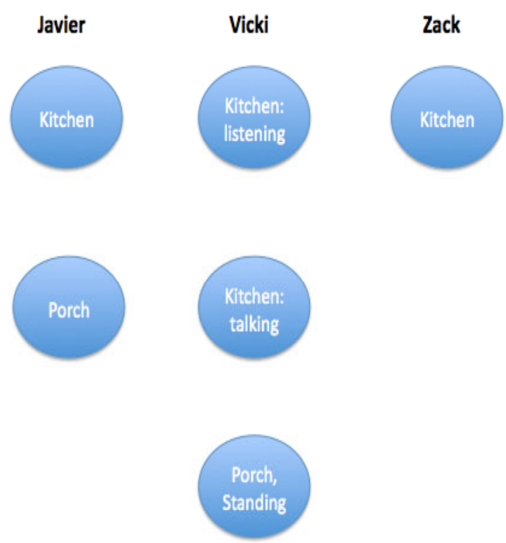


Fig. 2. The control map used in early tests

The inter-actor moves the control away from its neutral position above one of the marks, and the corresponding gesture is smoothly carried out. Associated with each mark, up to nine gestures can be mapped onto the four available directions of motion parallel to the table-top and directly up from the table (each lateral direction can contain up to two gestures). Diagonal motion results in a blend of the corresponding discrete gestures.

For instance, the dot at Javier: Kitchen controlled four gestures. Moving the marker away ("north") caused Javier to lean forward on the kitchen table. Moving to the east (right) generated a "chest bump" - an assertive thrusting outward of Javier's chest. Moving south caused Javier to cross his arms. Further movement southward produced an opening of the arms, in a palms-up gesture. Movement westward caused Javier to rotate toward his right, and to look "flirty"; i. e. downward, with eyes upward.

In each case, the inter-actor's head motions (tracked by means of markers on a baseball cap) are added to those based on the pose control system. Thus a gesture which turns the head to the left, for instance, can still be modified with additional left/right or up/down movement of the inter-actor's head.

3.2 Scenario-Specific Gestures

The limitations of this technique are obvious: only those gestures that have been prepared for a given scenario are available. However, there are several advantages as well. The motion capture suit of TeachMe requires that the inter-actor have a rich repertoire of 'body language', with which to express the personalities of all the characters in the story. The new technique requires good control of timing but relaxes the requirement for specific body movement skills. It also frees up the inter-actors' hands for other purposes, when required.

Two of those purposes are to move the avatar's lips when speaking, and to control facial emotions. An Ergodex keyboard [4] or similar device is used to move the lips, and to select facial expressions.

3.3 Designing the 3d Characters

The team worked with teenage consultants to gather images of prototypes for the characters in the story. The inter-actors posed the various postures and gestures that would be needed, and photographs were provided to the 3d modelers.

The team generated a specific list of emotions that the faces of the characters would need to convey. These included:

confused
interested
excited/flirty
pleased/happy
concerned
disgusted
'whatever'
displeased/hurt

An obvious question would be why no expression of anger was included. This choice was made because, based on the context of the story, there is no expectation that the onscreen characters will need to express anger. The participant may well experience that emotion or others, but the available expressions coupled with appropriate tone of voice and dialog could adequately deliver such emotions.

The acting team produced these facial expressions and then developed consensus as to which expression best reflected what they wanted on the characters. Photographs of these inter-actor-generated emotional faces were then provided to the artists.

3.4 Test Configuration

The next round of in-lab tests were set up as depicted in Figure 3 below.

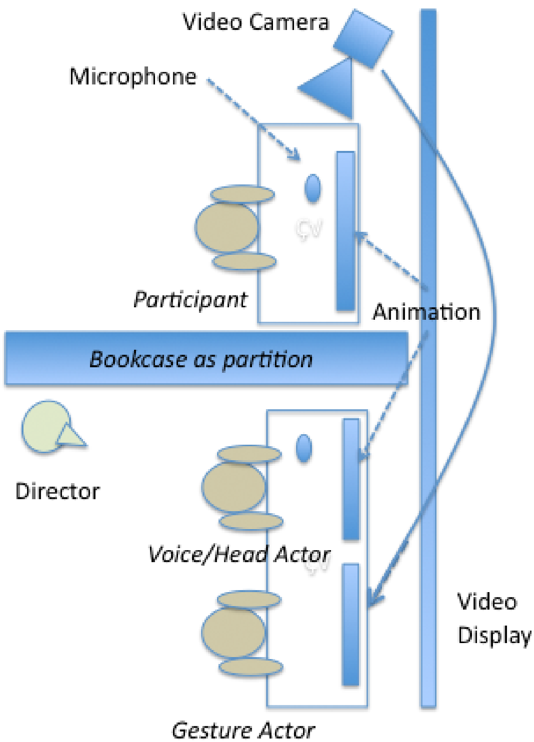


Fig. 3. Setup for Laboratory Tests of 3d System

During early trials, two inter-actors shared control of the avatars. One (wearing the motion capture hat) controlled the head's movement, spoke the lines and selected the facial expressions of the avatar. The second controlled the gestures. The show director, sitting behind the inter-actors, provided narration and changed the scene when appropriate.

During later trials, one inter-actor performed all the tasks previously handled by two. The second inter-actor (now sitting where the director is marked in Figure 3) was in charge of narration, scene changes and also with providing feedback to the participant. The show director remained close by, but took no active role in the scenario while it was running.

3.5 A Feedback System

In the original design of the scenario, it was intended that the participant should be given continuous feedback by an on-screen display of some kind. The first version was a display with two columns whose height represented 'Strength' and 'Coolness', with a horizontal bar across the top. The participant was instructed to try to keep the two parameters in balance.

As pilot experiments continued, it became clear that it was quite difficult for the narrator to reliably assign scores to the participant's actions. Also the scoring was not providing feedback specific to the two peer resistance skills being targeted by game play. The feedback was modified to target two peer resistance skills, resistance and avoidance. Consistent with the conceptual framework guiding game design ('Communication, Competence Model')[5], the points earned for resistance and avoidance can be increased by demonstrating each behavior in a “cool” way.

It will be part of the research component of the project to determine if the narrator will be able to assign scores in a reliable and reproducible manner. In this paper our principal concern is with the cognitive loading and overall task performance of the inter-actor and narrator.

4 Interactive Performance Issues

The inter-actors involved in this project have several years of experience in live interactive performance. They were not daunted by the challenge of playing five characters in rotation. The new control interface took some getting used to, but the inter-actors report that they feel that they have good control over the avatars.

One of the advantages of the five dimensional interface is that, if an inter-actor is unsure what gesture a given control movement will produce, they can make a small movement and observe the beginnings of the gesture, then retract if necessary. To the participant, the movement looks like natural 'fidgeting' behavior of the avatar character.

A more significant challenge is to make sure that the participant encounters all of the experiences planned for in the experimental treatment plan. A paper check-list is currently being used. The narrator is responsible for tracking whether the essential story points come up, and for prompting the inter-actor if necessary before a scene change, to include all the expected experiences.

We are using an inter-actor as the narrator because the narrator is playing an essential role in the dramatic process. For consistency in the evaluation process, this person will play the role of narrator/evaluator throughout the school experiment.

5 Experiments to Be Conducted

At press time, the actual experiments are scheduled to take place in April 2011 in the after-school program held at a public middle school in Orange County, Florida. The inter-actors will be working from the Institute for Simulation and Training at UCF, approximately fifteen miles from the school. The outcomes of the experiments with respect to interactive performance will be presented at the HCII 2011 conference. The outcomes with respect to the training goals of the project will be presented at conferences and published in nursing and interdisciplinary journals that focus on health promotion.

Acknowledgements. Funding for this project was provided by: College of Nursing and the Institute for Simulation and Training at the University of Central Florida, and the National Institute for Nursing Research, NINR (R15NR012189-01).

References

1. Wirth, J.: Interactive Acting: Improvisation, and Interacting for Audience Participatory Theatre. Fall Creek Press, Fall Creek Oregon (1994)
2. Milgram, P., Kishino, A.F.: Taxonomy of Mixed Reality Visual Displays. IEICE Transactions on Information and Systems E77-D(12), 1321–1329 (1994)
3. Dieker, L., Hines, M., Hughes, C.E., Smith, E.: Implications of Mixed Reality and Simulation Technologies on Special Education and Teacher Preparation. In: Focus on Exceptional Children, vol. 40, pp. 1–20 (2008)
4. Ergodex controller, `http://www.ergodex.com`
5. Spitzberg, B., Hecht, M.: A component model of relational competence. Human Communication Research 10(4), 575–599 (1984)

Emotions and Telerebabilitation: Pilot Clinical Trials for Virtual Telerebabilitation Application Using Haptic Device and Its Impact on Post Stroke Patients' Mood and Motivation

Shih-Ching Yeh[1], Margaret McLaughlin[2], Yujung Nam[2], Scott Sanders[2], Chienyen Chang[2], Bonnie Kennedy[2], Sheryl Flynn[2], Belinda Lange[2], Lei Li[3], Shu-ya Chen[2], Maureen Whitford[2], Carolee Winstein[2], Younbo Jung[4], and Albert Rizzo[2]

[1] National Central University, Taiwan
[2] University of Southern California, USA
[3] Microsoft, USA
[4] Nanyang Technology University, Singapore
shihchiy@csie.ncu.edu.tw

Abstract. We describe a pilot clinical trial with a flexible telerehabilitation platform that allows a therapist to remotely monitor the exercise regimen and progress of a patient who previously suffered from a stroke. We developed virtual game environments which were host to a progressive set of training tasks from precise fine motor movements to reaching movements that involve full arm and shoulder activity. Concurrently, the therapist monitored the progress of the patient through a video channel. Assessment of psychosocial variables show that negative feelings (confusion, $t(13)=2.54$, $p<.05$, depression $t(13)=2.58$, $p<.05$, and tension, $t(13)=2$, $p<.1$) were significantly lessened after the game play. Patients' overall satisfaction with the telerehabilitation system was positively correlated with the feeling of co-presence of the therapist, $r(8)=.770$, $p<.005$. Patients felt less efficacious in continuing therapy after participating in the telerehabilitation game compared to their reported perseverance self-efficacy before the game, $t(5)=2.71$, $p<.05$ and showed decreased willingness to persist in therapy regardless of fatigue after the game play, $t(5)=2.67$, $p<.05$. However, when patients' pretest mood scores were taken into account, this trend was reversed. Patients' active mood before the game was positively correlated with their willingness to persist in the therapy after the game, $r(14)=.699$, $p<.005$. Telerehabilitation significantly enhanced stroke patients' psychological states.

Keywords: Virtual reality, stroke rehabilitation, telerehabilitation, haptics.

1 Introduction/Background

For a number of years a large interdisciplinary team of researchers at the University of Southern California has been engaged in a series of studies of virtual environments

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 119–128, 2011.

and games for recovery of motor function following a stroke. Our goals have been to create rehabilitation scenarios and exercises that maximize therapeutic movement with cortical reorganization goals in mind; motivate patients to persist in training using gaming features; provide immediate feedback; are easy for therapists to administer and monitor; provide useful data in repeated-measures designs; and result in improved real-world function. Some of our efforts in this regard are reported in [1-4]. We have also turned our attention to extending this work into distributed environments where patient and therapist are not co-located. Well-defined and standardized outcomes used in telerehabilitation studies indicate that progress in addressing motor deficits may be addressed by remotely delivered therapeutic regimens [5] but the emotional and cognitive impact on patients may not be addressed. It is well documented that post-stroke patients suffer wide ranging psychological and cognitive disorders induced by cognitive impairment from damage to one side of their body or other limitations.

Below we report on a pilot trial investigating some factors influencing the satisfaction of therapists and patients with a telerehabilitation system and the impact of telerehabilitation game play on patient mood states and willingness to persist in therapy.

The primary argument for using games or game-like exercises for motor rehabiitation is that they may be sufficiently enjoyable to motivate patients to persist where other less engaging methods result in a substantial loss of interest in continuing training. Design choices made in games and the virtual environments in which they are played may influence enjoyment through encouraging a sense of presence, the feeling of "being there" when the player becomes totally absorbed and awareness of the physical place of game play is lost. In telerehabilitation, the critical element may be the sense of co-presence. A satisfactory experience will require that both parties experience themselves as co-located and mutually aware of, responsive to, and responsible to one another. Based on the foregoing we expect that the telerehabilitation experience will create elevated mood states and that the extent to which patients and therapists experience a sense of co-presence will be associated with satisfaction with the telerehabilitation system and willingness to persist in therapy. Thus

- H1: Mean scores on negative feelings after playing the telerehabilitation games will be lower than the scores before the game play.
- H1a: Patients' scores on feeling confused will be lower after playing the telerehabilitation games than before the game play
- H1b: Patients' scores on feeling depressed will be lower after playing the telerehabilitation games than before the game play
- H1c: Patients' scores on feeling tension will be lower after playing the telerehabilitation games than before the game play
- H2: Patients' overall satisfaction with the telerehabilitation system is positively correlated with the level of presence of therapists during the game play.
- H3: Patients will demonstrate increased willingness to persist in therapy after the game play.
- H3a: Patients will demonstrate increased willingness to adhere to daily therapy regimen after the telerehabilitation game play.

- H3b: Patients will demonstrate increased willingness to persist in therapy regardless of fatigue after the telerehabilitation game play.
- H4: Patient's active mood before the game play is positively correlated with their willingness to persist in the therapy after the game play.

2 Materials and Methods

Approval for the study was obtained from the Institutional Review Board at the University of Southern California. All participants were provided with an information sheet describing the purposes and requirements of the study and written consent was obtained. Participants were recruited for either a patient or a therapist role. Inclusion criteria for the patient role were impairment of motor coordination in a upper extremity following stroke, traumatic brain injury, spinal cord injury, or amputation. Exclusion criteria were sensory, cognitive or linguistic limitations that would be make it difficult for them to respond to a computer interface, follow instructions, or complete questionnaires orally in response to an English-speaking interviewer. The inclusion criterion for the therapist role was that the individual was currently practicing occupational or physical therapy in California.

Patient volunteers were recruited at Precision Rehabilitation, Rancho Los Amigos Rehabilitation Clinic and other rehabilitation clinics in Southern California. Clinic managers and staff therapists were provided with flyers to raise patient awareness of the study and provide potential patient volunteers with the investigators' contact information if they were interested in finding out more about the study. Therapist volunteers were recruited from Precision Rehabilitation, Rancho Los Amigos Rehabilitation Clinic and other rehabilitation clinics in Southern California. Recruitment took place through clinic managers who were provided with flyers and a form letter email describing the study and the investigators' contact information.

2.1 Experimental Protocol

All telerehabilitation sessions were held at the facilities of Precision Rehabilitation in Long Beach, California. This outpatient facility is focused primarily on rehabilitation of individuals with spinal cord injury, traumatic brain injury, stroke and amputation. Therapist/patient pairs were taken into separate rooms. A third room was used for administrative tasks and questionnaire administration .Each of these rooms had a door which remained closed during the questionnaire administration and computer game sessions so that study activities would not be accessible or visible to others at the center who were not directly involved.

Each participant had the study explained to him or her and each was provided with the opportunity to ask questions. Following informed consent, the patient and therapist separately completed a series of questionnaires (see Measures, below). The therapist was provided with an instruction sheet for the technical aspects of the telerehabilitation session and a script to engage the patient in the therapeutic tasks. Both participants were informed that they would communicate with one another over an Internet connection between the rooms. The therapist's role was to guide the patient through the setup of the systems and then talk him/her through three computer games designed to provide motor rehabilitation exercises for the upper extremity. Following completion of the tasks, the patients and therapists were asked to complete a usability questionnaire answering a series of open-ended and closed-ended questions and a post-test questionnaire about willingness to persist in training. Study personnel able to handle technical issues with the networking components or assist the patient with equipment use were available at all times.

2.2 System Description

The telerehabilitation system was composed of two subsystems: a motor rehabilitation system and a tele-communication system.

Motor Rehabilitation System

Task 1: Plane Flying. This task was designed for practice in wrist pronation/supination. In the virtual environment; there was an airplane moving forward with a constant speed and participants had to manipulate the airplane up-down or left-right and roll the airplane via wrist pronation/supination in order to fly through a sequence of hollow rectangular barriers set with different rolling angles. To interact with the virtual environment, the user employed an Omni haptic device. The Omni provided force feedback as an indication to the participants that the airplane had hit the rectangular barrier. To come up with various difficulty levels so that participants were able to repeatedly challenge the limits of their current motor capability, several parameters were chosen and combined in different ways. First, the size of the rectangular barriers (width and height) was a parameter that determined how accurate and stable the movement was required to be to position the airplane appropriately without hitting the edges of the barrier. The smaller the size of the barrier, the greater the accuracy and stability of wrist movement that was required. Second, the "rolling angle" of the barrier was set to direct the offset of wrist rotation from a neutral position. The larger the rolling angle of the barrier, the more rotation of the wrist was required. Third, the "speed" of the moving airplane was a parameter determining how soon in the user's forward motion the airplane approached the next rectangular barrier. A higher moving speed of the airplane meant that participants had a shorter reaction time to respond. Three difficulty levels were designed with the combination of the parameters above. To measure the performance of participants, a rate of passing through the barriers was defined as the ratio of the number of successful passing trials to the number of total trials.

Task 2: Water Pouring. This task was designed for the practice of wrist pronation/supination. In the virtual environment, participants manipulated a cup via wrist pronation/supination in order to fill it with water from a source water tank and pour the water into a target water tank. The interaction between participants and the virtual environment was achieved by a camera-LED tracking system. To define various difficulty levels so that participants were able to repeatedly challenge the limits of their current motor capability, several parameters were selected. First, the "size" of the source and target water tanks was a parameter to determine how accurate and stable the movement would have to be in order to fill or pour water without loss. The larger the size of the water tank, the less accuracy of movement required to fill or pour water. Second, a threshold with respect to the rolling angle offset from the neutral pose of the wrist was set to determine if the water was able to flow into/out of the cup. Specifically, the rolling angle made by the wrist should exceed the threshold in order to activate the water flow. A higher threshold required a larger rolling angle of the wrist to complete the task. With the combination of parameters mentioned above, three difficulty levels were designed. The total time to complete each task was recorded as a measure to evaluate participants' performance. Performance score was further defined as the summation of the reverse of total time at each level, multiplied by a weighting factor for each difficulty level. A higher score meant a better performance.

Task 3: Ball Dropping. This task was designed for the practice of hand opening. In the virtual environment, participants manipulated a virtual hand with a ball in it. Their task was to reach an object hanging above a grid and drop the ball via hand opening as rapidly as possible. A 5-DT data glove and the camera-LED tracking system were used to sense the hand opening and hand movement respectively. A difficulty level could be determined by the degree of finger extension which defined hand opening and taking into account the initial calibration for each participant. (However, in this pilot study we set the degree of finger extension as a constant). To measure the performance of participants, several measures were defined. First, movement efficiency, defined as the length ratio of the actual moving path over the shortest possible path was used to evaluate the efficiency of participants' moving paths. A smaller value of moving efficiency represented greater moving efficiency. Second, the time (move time) spent in moving the virtual hand from the start position to the top of the target object was recorded. A lower value of move time represented greater mobility. Third, the time (hand open time) spent in opening the hand was recorded as an indication of the dexterity of hand opening. A lower value of hand open time represented better performance.

Tele-communication System. The tele-communication system was designed with the use of peer-to peer network technology as a flexible platform that can deploy applications which require high-speed transmission of delay-sensitive media streams, such as audio, video, and miscellaneous data. Beyond the transmission of media streams, three functionalities were further developed to satisfy various objectives from the perspectives of therapy, usability and human perception. First, SkypeTM was embedded into our framework to enable oral and visual communication (with the use of webcams) between the patient and the therapist; thus, the therapist was able to remotely monitor the exercise regimen and progress of a patient and direct the patient through a complete practice session using interactive VR tasks as in clinics. Second, while the patient focused on interaction with the game environment, a view-synchronization mechanism was made so that the view of the game environment as seen by the patient was transmitted to the therapist's screen. Thus, in addition to watching the live video stream of the patient's behavior through a second camera focused on the patient's arm, the therapist could also see exactly what the patient saw on screen. The system was designed to enhance the sense of presence in the virtual environment and co-presence between the patient and the therapist who were located remotely. Finally, a remote database system was constructed to store various measures of the patient's performance, such as motor data, task completion time, task completion rate, etc., which were transmitted immediately and continuously via DSL.

The therapist was able to retrieve the data of the patient's performance at a later time for further analysis.

(a) (b) (c)

Fig. 1. (a) Plane Flying task. (b) Water Pouring task. (c) Ball Dropping task.

2.3 Measures

Data collected from patients included a demographic and computer use questionnaire; the mood as measured by the POMS-A; experience of "presence" in the telerehabilitation environment, willingness to persist with therapy, and a telerehabilitation usability questionnaire. To therapists we administered a demographic and computer use questionnaire, a presence measure, and a usability questionnaire.

Mood was operationalzed by modifying the Brunel Mood Scale (POMS-A). Composite indices of three mood items, (1) negative, (2) confused and (3) active mood, and one item measuring depression were drawn from the relevant POMS-A items. Evaluation of negative mood was created combining the eleven 5-point scale negative attitude items (Chronbach's alpha = .953). This includes the following individual mood items; feeling downhearted, annoyed, exhausted, bitter, unhappy, anxious, worried, miserable, angry, tired and bad tempered. Confusion was a composite index combining the following four items from the POMS-A; feeling confused, mixed-up, sleepy, and uncertain (Chronbach's alpha=.819). Evaluation of active mood was an index composed of the following four items; feeling lively, energetic, active, and alert (Chronbach's alpha = .952).

Telerehabilation usability for the patient was an composite index of thirty-seven 7-point scale items, including "The telerehab system was easy to use;" "I felt comfortable using this system;" "The instructions helped me complete this task;" "I learned to do the movements quickly with videoconferencing;" and "Overall I am satisfied with the amount of time it took to do the exercises" (Chronbach's alpha = .917). Telerehabilitation usability for the therapist was composed of twenty-six 7-point scale items (Appendix A) including; "I am satisfied with the feedback system;" "The feedback window easily allowed me to see changes in the patient's performance;" "I would recommend the telerehabilitation system to other people;" "Overall, I have the impression that the telerehabilitation system will increase the scope of patient care/treatment;" "Overall, I have the impression that the telerehabilitation system will increase patients' accessibility to clinicians" (Chronbach's alpha = .909).

Willingness to persist in therapy was measured by two 7-point scale items; "I believe I can do therapy every day that it is scheduled;" and "I believe that I can do therapy no matter how tired I may feel."

Presence in the telerehabilitation environment for patients was measured with five 7-point items; "I felt like I was with the therapist;" "There were times I felt like the therapist was with me in the same room;" "I felt like the therapist and I worked well together;" "I felt like the exercises using the telerehabilitation system were the same as working with the therapist face-to-face"; "I felt that there was another person working with me on the exercises" (Chronbach's alpha=.815). Presence for therapists was similarly measured by five 7-point scale items such as the following; "To what extent, if at all, did you have a sense of being with the patient?"; "To what extent were you and the patient in harmony during the course of the exercises?"; and "Overall, rate the degree to which you have a sense that there was another person interacting with you during the exercises" (Chronbach's alpha = .882).

3 Results

Paired-samples t tests were conducted to evaluate whether a patient's negative mood states were lessened after playing the telerehabilitation games. The assessment of psychosocial variables showed that negative feelings were significantly lessened after the game play. The mean score for feeling confused was significantly less after game play (M = 4.50, SD = .76) than the scores before the game play (M = 6.58, SD = 3.07), t (13) = 2.54, p < .05, (*Ns* for all were 14). Participants reported feeling less depressed after game play (M = 5.17, SD = 1.49) than before playing the game as well (M = 6.67, SD = 3.15), t (13) = 2.58, p < .05. The mean score on feeling tension was also significantly lowered after playing the games (M = 6.00, SD = 2.08) than before (M = 7.08, SD = 3.00), t (13) = 2, p < .1. Overall, we found that participants felt less confused, less depressed and more relaxed after playing the games, supporting hypotheses 1(a)(b)(c).

A correlation coefficient was computed between the patients' overall satisfaction with the telerehabilitation system and the therapists' level of co-presence during the game play. The result showed that the correlation of the participant's overall satisfaction and the therapists' feeling of presence was statistically significant, r (8) = .770, p<.05, N=8. Overall, therapists whose patients rated the telerehabilitation system more useful and satisfactory were more likely to feel higher co-presence during the game play and vice versa, supporting H2.

Paired sample t tests compared the pre- and post-test scores of patients' willingness to persist in therapy. The result of the mean comparison between pre- and post-test data did not provide support for the hypotheses on the positive impact of the game play on patients' level of perseverance. Patients reported that they felt significantly less capable of following through daily therapy regimen after the telerehabilitation game play, t(5)=2.71, p<.05, N=6 (M=6.83, SD=.41 in the pre-test, M=6, SD=.63 in the post-test). Similarly, they felt significantly less capable of persevering in the therapy regardless of fatigue after the telerehabilitation game compared to before the game play, t(5)=2.67, p<.05, N=6 (M=6.16, *S*D=.75 in the pretest, M=5.35, SD=.50 in the post test). The results show that patients felt less efficacious in continuing therapy after participating in the telerehabilitation game compared to their normal level of perseverance.

A correlation analysis supported hypothesis 3. An active mood before the experiment was positively correlated with patients' willingness to persist in the therapy after the game, r (14) =.699, p < .005, N=14. The results suggest that if patients report that they feel active and motivated, they tend to state that they are willing to continue in the therapy after the game play regardless of their experience with the telerehabilitation games.

4 Discussion

This pilot clinical trial examined whether a telerehabilitation platform with a built-in video communication component between a therapist and a post-stroke patient would be effective in improving patients' psychological states and motivation to persist in training. The telerehabilitation system was host to a set of virtual games and a video communication module that ran concurrently with the game platform. The three virtual rehabilitation games contained full arm and shoulder training activities with a focus on pronation and supination and hand opening and closing. The difficulty levels and the progress in gameplay was monitored and manipulated through a live video chat during the exercise. Post-stroke patient volunteers with impairment of motor coordination in the upper body and staff therapists who monitored and guided the patients through virtual game play participated in the pre-post design experiment. The task involved completing three virtual tasks: 1) The plane flying task involved manipulating a virtual airplane to fly through a sequence of rectangular barriers by pronating and supinating the wrist using an Omni haptic device; 2) The water pouring task was designed for improving motor capability by filling a virtual water cup from the source and pouring it into the target tank. A camera-LED tracking device tracked the wrist pronation and supination mapped onto the virtual objects; 3) The ball dropping task involved a hand opening and closing motion. Patients reached a virtual object above a grid and dropped it at the target by opening their hands. A 5-DT data glove with the camera-LED tracking system traced the hand opening and closing motion.

The results showed that the telerehabilitation games for recovery of motor function were effective in producing psychological gains and a sense of co-presence. However, these advantages in changing emotions and mood were closely related to their psychosocial states before the game play. The therapist's feelings and level of co-presence was shown to affect the patient's overall satisfaction with the telerehabilitation system.

Patients' negative mood states were significantly lowered after playing the telerehabilitation virtual games. All three aspects of negative emotions including confusion, depression and tension, assessed by related psychosocial variables, were lessened after completing three virtual telerehabilitation tasks. Our previous work has reported functional advantages of the telerehabilitation platform in improving therapeutic movement with cortical reorganization and regaining motor function of an upper extremity among post-stroke patients. This finding demonstrates that telerehabilitation exercises using video game features such as an immediate reward and encouragement can positively impact the patient's psychological states as well. Mood disorders such as anxiety and depression followed by physical disability and intellectual impairment from stroke are prevalent among post-stroke patients [6]. Neuropsychological studies show that the prevalence of negative emotions after the stroke are often associated with negative impact on motor, cognitive and intellectual recovery such as lack of motivation in participating in therapy, slow progress in recovery, longer stays in the hospital, less engagement with leisurely activities and even lower survival rate [7]. Given a strong association between less positive moods and the negative impact on recovery from stroke, it is important to address how the telerehabilitation system can enhance emotional states and sense of wellbeing of patients.

The study also showed that participants whose therapists felt a higher sense of co-presence during the game play reported higher overall satisfaction with the telerehabilitation game play experience. When therapists felt more immersed in the telerehabilitation experience and fully engaged with the presence of the patients, patients' general sense of acceptance of the remote rehabilitation session and positive evaluation of the overall experience were higher. The quality and experience of participating in the telerehabilitation exercise should be comparable to the clinic-based therapy where physical therapists are physically present with the patient [8]. An experience of spatial co-location, mutual comprehension, and emotional closeness in a virtual environment is crucial in creating a higher sense of co-presence [9]. The telerehabilitation platform with a built-in system design that allows patients and therapists to see and communicate with each other in real-time enhanced the level of engagement and communication between patients and therapists. Therapists could closely evaluate the patient's movement progress, performance updates and emotional states and manipulate the difficulty levels and pace of the game play accordingly. It is important to recreate a comparable sense of co-presence between therapist and patient in a remote telerehabilitation session as in onsite training since it plays a large role in the patient's immediate sense of satisfaction upon completing tasks. Despite the technological advancement and sophisticated telerehabilitation platform, a good relationship between the patient and the therapist and their sense of working together still contributes significantly to the rehabilitation process.

5 Conclusions

We hypothesized that patients would demonstrate increased willingness to persist in therapy after the telerehabilitation game play. However, patients felt significantly less willing to adhere to their daily therapy after the telerehabilitation game play. They also reported decreased willingness to persist in therapy regardless of fatigue after the participation. The telerehabilitation training seemed to have a negative impact on patient's level of perseverance and motivation from the initial analysis. However, the trend was reversed when the patient's baseline mood was considered in the interpretation. When patients felt active and motivated in the pretest, their willingness to persist in therapy was more likely to increase after the game play. Many studies have documented a variety of health consequences resulting from lower health efficacy levels and lack of motivation [10]. However, it has not been widely investigated whether level of stimulation and elated mood in a virtual telerehabilitation setting is related to patient's willingness to persist in therapy afterward. This result confirms the previously mentioned relationship between mood states and their potential impact on the effectiveness of telerehabilitation training. The finding demonstrates that lack of motivation and negative moods could be related to a failure to adapt to the exercise regimen in a virtual telerehabilitation training session.

References

1. Jung, Y., Yeh, S., McLaughlin, M., Rizzo, A., Winstein, C.: Three-dimensional game environments for recovery from stroke. In: Ritterfeld, U., Cody, M., Vorderer, P. (eds.) Serious Games: Mechanisms and Effects, Routledge.Natl ons (2009)
2. McLaughlin, M.L.: Simulating the sense of touch in virtual environments: Applications in the health sciences. In: Messaris, P., Humphreys, L. (eds.) Digital Media: Transformations in Human Communication. Peter Lang Publishers (2005)
3. Yeh, S.-C., Stewart, J., McLaughlin, M., Parsons, T.D., Winstein, C.J., Rizzo, A.A.: Evaluation approach for post-stroke rehabilitation via virtual reality aided motor training. In: Dainoff, M.J. (ed.) HCII 2007 and EHAWC 2007. LNCS, vol. 4566, pp. 378–387. Springer, Heidelberg (2007)
4. Stewart, J.C., Yeh, S., Jung, Y., Yoon, H., Whitford, M., Chen, S., Li, L., McLaughlin, M., Rizzo, A., Winstein, C.: Intervention to enhance skilled arm and hand movements after stroke: A feasibility study using a new virtual reality system. Journal of NeuroEngineering and Rehabilitation 4, 21 (2007), doi:10.1186/1743-0003-4-21
5. Holden, M.K., Dyar, T.A., Dayan-Cimadoro, L.: Telerehabilitation using a virtual environment improves upper extremity function in patients with stroke. Transactions on Neural Systems and Rehabilitation Engineering 15(1), 36–42 (2007)
6. Berg, A., Lonnqvist, J., Palomaki, H., Kaste, M.: Assessment of depression after stroke: a comparison of different screening instruments. Stroke 40(2), 523 (2009)
7. Spencer, K.A., Tompkins, C.A., Schulz, R.: Assessment of depression in patients with brain pathology: the case of stroke. Psychological Bulletin 122, 132–152 (1997)
8. Lewis, J.A., Boian, R., Burdea, G., Deutsch, J.E.: Remote console for virtual telerehabilitation. In: Proc. MMVR 2005, Long Beach, CA, January 2005, pp. 294–300 (2005)
9. Biocca, F., Burgoon, J., Harms, C., Stoner, M.: Criteria and scope conditions for a theory and measure of social presence. Media Interface and Network, E. Lansing (2001)
10. Miller, W.R., Rollnick, S.: Motivational Interviewing: Preparing People for Change, 2nd edn. Guilford Press, New York (2002)

An Interactive Multimedia System for Parkinson's Patient Rehabilitation

Wenhui Yu, Catherine Vuong, and Todd Ingalls

School of Art, Media and Engineering,
Arizona State University
Tempe, AZ 85287
{Wenhui.Yu,Catherine.Vuong,Todd.Ingalls}@asu.edu

Abstract. This paper describes a novel real-time Multimedia Rehabilitation Environment for the rehabilitation of patients with Parkinson's Disease (PD). The system integrates two well known physical therapy techniques, multimodal sensory cueing and BIG protocol, with visual and auditory feedback to created an engaging mediated environment. The environment has been designed to fulfill the both the needs of the physical therapist and the patient.

Keywords: Parkinson's Disease, Physical Therapy, Mediated Rehabilitation, Sensory Cueing, Multimodal Feedback, Virtual Environment.

1 Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disease that affects nearly 1% of the elderly population [11]. It is characterized by akineisa (impaired initiation of movement), bradykinesia (reduced amplitude and velocity of movement), tremor, rigidity, and postural instability [6][18].The symptom are believed to be caused by the degeneration of the dopaminergic nigrostriatal pathway, a major pathway supplying dopamine to the striatum [6].

Currently, the standard treatment is L-3, 4-dihydroxyphenylalanine (Levodopa or L-DOPA) therapy. L-Dopa has been shown to improve the motor symptoms and life expectancy after the onset of PD. However, these benefits are not permanent; typically, within five years the gains begin to wear off and produce negative side effects [5]. Neurosurgery, which destroys a small portion of the brain, and deep brain stimulation are both treatment possibilities if L-Dopa becomes ineffective; but these are expensive and contain significant risk [6].

Given the side affects of the medication, physical therapy can be utilized to delay the need for drug interventions or used simultaneously to reduce the side effects. Our team has designed and developed a Parkinson's Multimedia Rehabilitation System that combines physical therapy techniques, sensory cueing [2][3][9][12][21][20][21] and the BIG protocol [3][7][8], with visual and auditory feedback to create an engaging mediated environment.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 129–137, 2011.

2 Background

Sensory cueing, generally in reference to visual and auditory cues, is the use of external stimuli to facilitate movement. For example, of visual cues can be markers or lines on the floor to encourage PD patients to take bigger steps. This has been shown to increase stride length and stride length regulation in PD patients [2][19][20][21]. Auditory cues, such as music or a metronome, has been shown to increase cadence (rhythm) and velocity in PD patients [10][17][19].

The BIG protocol is the use of large amplitude movements to encourage patients to move “bigger” [3][7][8]. Although this technique encourages larger than normal movement it is believed to have carryover effects that increase the amplitude of general movement in PD patients’.

Multimedia rehabilitation systems have started to become more widespread with examples in stroke, cognitive rehabilitation, balance and, telerehabiliation [13][14][16]. One approach has been through the use of traditional Virtual Reality (VR) and Virtual Environments (VE) for motor training [13][16][23]. Similar approaches have long been suggested for PD and some earlier studies were promising [1][4]; however, recent literature points to some unforeseen problems. In recent literature, while using VE that replicates activities of daily (i.e. walking in a kitchen, supermarket, or classroom), using a full head mounted display, PD patients experienced hallucinations (i.e. children sitting in the desks when there were no children in VE) during the “off” phase of L-Dopa therapy. None of the hallucinations were the same and not all PD patients experienced it. This should not present an issue as we are not replicating activities of daily living or using a full head view display.

In this paper we will describe the design and development of a real-time Parkinson’s Multimedia Rehabilitation System. We will describe the integration of sensory cueing techniques and the BIG protocol with visual and auditory feedback used to create an engaging environment for rehabilitation.

3 User Needs and Strategies

In order to create a system that meets both the needs of the physical therapist and of the PD patients we have identified a series of needs which are described below.

3.1 Physical Therapist Needs

After discussion with a PD occupational physical therapist and literature review, we have broken the requirements into three main categories that we believe fulfill the needs of the therapist:

3.1.1 Usable in a Clinical Space

Our current system works within our laboratory setting, however, we hope that future revisions of our system can be implemented in a clinical setting. We are trying to implement a system that is cost effective and space efficient. We therefore be limited the number of computers, projectors, projection screens, and active space in our environment with this need in mind.

Treats the Symptoms of PD. In the development of the system, we targeted the symptoms associated with akinesia (impaired initiation of movement) and bradykinesia (reduced amplitude and velocity of movement). These symptoms not only result in hypokinetic (reduced movement) movement characteristic to PD but also contribute gait irregularities such as difficulty in stride length regulation, regular cadence (rhythm), and low velocity [6][19]. As a result the following need were identified:

- BIG and Faster: The system was designed to ameliorate the symptoms related to bradykinesia by encouraging larger and faster movements. The system tasks integrates the amplitude-based whole body exercise tasks and speed to promote the patient's ability to reach and step as far and fast as possible.
- Accuracy and Timing: The system was designed to alleviate the kinetic symptoms associated with akinesia and bradykinesia by increasing accuracy and timing of a patients' reaching and stepping. To do this the accuracy measurements were made flexible and reasonably tolerant to meet the needs of the patient. Focus was placed on timing, faster reaction and fluidity of movement.
- Repetitive and Variable: The tasks presented to patients in the system were designed to be either repetitive or variable. Repetitive tasks (a single task repeated over and over) can help patients learn new movement patterns. Variable tasks (multiple task mixed together) can allow patients to learn how to transition from one movement pattern to another and provide variation in the task.

3.2 Patient Needs

The needs of patients were also taken into consideration. To meet the needs of the patient the system was designed to be engaging, immersive, and adaptable.

Engaging and Immersive. The effectiveness of the environment is based upon the active participation of the patient. During a long-term repetitive exercise session, patients can easily become physically tired and mentally bored. Thus, it is necessary to design an engaging and immersive environment, which can help patients remain attentive and motivated.

Adaptable. Given that patients may get frustrated when failing to reach/step the target the system must be adaptable. It must be able to push the patients limits, but adaptable so that patients do not become frustrated with the task.

4 System Design

In this system, the location of the patient's hands, feet, and torso are mapped directly to an avatar. Patients are asked to control this avatar to complete movement tasks in the game, projected on the screen. The environment integrates sensory cueing into game-like mediated instruction. Patients are given feedback based on accuracy and timing of movement performance.

Fig. 1. Displaying the sky scenario, in which the avatar stands on the cloud to reach falling stars. This environment provides several different scenarios, with different scenes (i.e. galaxy, sky, spaceship), sets of falling object objects (stars, rain, alien), and avatars, to engage patients.

This environment include 11 different simultaneous reaching/stepping patterns the patient can perform. They include: 1) RH, 2) LH, 3) RF, 4) LF, 5) RH/LH, 6) RH/RF, 7) LH/LF, 8) RH/LF, 9) LH/RF, 10) RH/LH/RF, 11) RH/LH/LF; where RH stands for right hand reach, LH for left hand reach, RF for right foot step and LF for left foot step. These patterns can either be performed individually (i.e. patient does 20 RH reaches) or in some combination (i.e. patient does 20 random RH, LH, RF, or LF reaches/steps). This flexibility, in determining the movement patterns, will allow the physical therapist to modify the tasks to either be repetitive or variable and adaptable (see Section 3.1.3 and Section 3.2.2) depending on the physical therapy needs of the patient.

The aim of the system is to motivate patients to follow these visual cues to reach or step on the expected positions in sync with music. In following sections we will describe the structure of the system, including: the composition, the integration of sensory cueing, feedback, and the BIG protocol, and how it addresses the needs of the physical therapist and the needs of the patient.

4.1 System Structure

The system is composed of an active space and the application. The active space is the physical space in which the patient interacts with the game. The application, is the software responsible for creating the virtual interactions between game and the patient. The application receives input data through a motion capture system and provides output through either sensory cueing or feedback via projector and speakers in the active space.

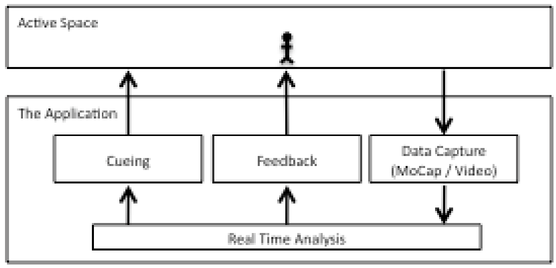


Fig. 2. High Level Design of the Environment

4.2 Data Capture

The system uses a 10 near-infrared camera Motion Analysis System to capture the location of the patient's hand, feet and torso. The data from the cameras are processed using EVaRT software (Motion Analysis Corp, Santa Rosa, CA) . The systems capture patient movement by tracking retro-reflective markers attached to the patient; each marker corresponds to anatomical labels in the EVaRT software. The overall system is run at a rate of 100 motion capture frames per second.

4.3 Sensory Cueing

In this subsection, we will detail how visual and auditory cueing are used to facilitate movement, particularly larger or BIG movement.

Visual Cueing. In the system, patients reach or step toward visual cues (as described in introduction to the section). The location of the visual cues are relative patients' arm/leg length with respect to the center of the screen.

There are several cues in this environment, which are divided into two categories: task and body cues. The first set of cues, task cues, help inform the patients of their task within the environment, this can be seen in Figure 3. These cues include: 1) starting cues, 2) target cues, 3) return cues and 4) moving cues.

The task cues are used as follows for each reach/step. A starting cue appears at the top of the screen, which informs the patient that the moving cue will appear there. Moving cues scroll downward from the top of the screen and pass over stationary semi-transparent target cues. The patient must reach (or step on) the target cue when the moving and target cue overlap. The moving cue attaches to the avatar and the avatar must return the moving the cue to the return cue. Once the avatar has reach the return cue, the moving cue disappears; thus completing the moving task.

This environment can be adapted to encourage larger movement by adjusting the location of the target cue, while the speed of the moving cue can be used to encourage faster movement.

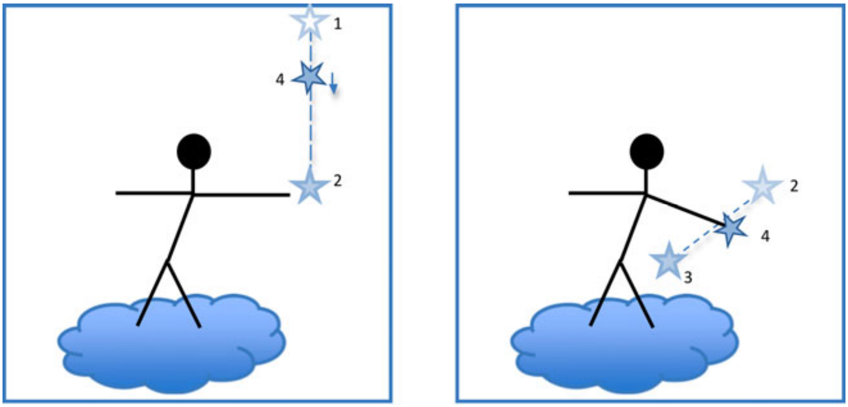


Fig. 3. Visual Task Cues for the Environment. In A, the Avatar is reaching toward the target cue (2). In B, the avatar has reached the target cue (2) and is now returning to the return cue (3). The cues in this figure are: (1) starting cue, identifies the horizontal location the target cue will appear, (2) target cue, the location the avatar is to reach, (3) return cue, the location the avatar is to return after reaching the target cue and (4) moving cue. At first, the moving cue scrolls down from the starting cue to the target cue. If avatar reaches the moving cue at the target cue position successfully then, the moving cue attaches to the hand/foot of the avatar. The moving cue will disappear once the avatar reaches the return cue. For clarity, the visualization in this figure have been reduce to a stick figure; an example of the actual visualization can be seen in Figure 1.

The second set of cues are body cues. Body cues are attached to each appendage and informs the patient which limb must or can move to reach/step on the target cue; this can be seen in Figure 4. This will be implemented using a traffic light color scheme. Green means the limb is involved in the reach and therefore must move to reach the target. Yellow means to reach the target the limb may move as a result of the reach. Red mean that this limb is not involved in the reach and therefore should not be moved; this will prevent patients from walking back and forth to reach targets.

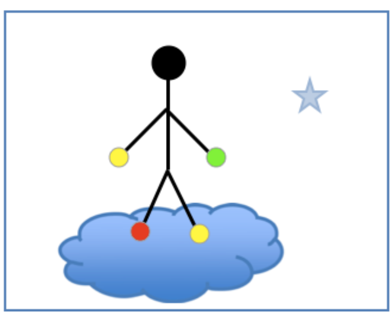


Fig. 4. Visual Body Cues for the Environment. In this example, the patient is asked to reach with the right hand to the falling star. To perform this task the patient has to move his/her right hand (in green). The patient should not move his/her left foot (in red). The movement of the left hand and right foot is not relevant to the task, thus it is also in yellow. For clarity, the visualization in this figure have been reduce to a stick figure; an example of the actual visualization can be seen in Figure 1.

Auditory Cueing. In this system, music with a strong tempo is played in the background to improve cadence and timing. The tempo of the music can be increased or decreased to encourage faster or slower movement. The slower tempo can be used if movement difficulty or fatigue arises; it can be also used to vary the movement task.

4.4 Sensory Feedback

In this system, feedback is given based on the successful completion of movement task (Knowledge of Results, KR) and on whether the movement is being performed correctly (Knowledge of Performance, KP).

As stated in Section 4.3.1, the task is to: 1) reach the moving cue as it passes over the target cue and 2) then to return the moving cue to the return cue. For both part 1 and 2 of the task, feedback is given based on whether the participant successfully completed the task or KR. In part 1, if the participant successfully reaches the moving cue as it passes over the target cue (within a given timing tolerance level), the moving cue shimmers, a happy melody is played, and the moving cue attaches to the limb of the avatar. Now, the shimmering moving cue can only be removed by returning it to the return cue, or second part of the task. In part 2, when the participant returns the shimmering moving cue to the return cue, the moving cue disappears. The timing tolerance level can be adjusted to improve timing, accuracy, or alleviate frustration.

In addition, patients are be given feedback based on whether they are performing the task correctly; this feedback has been integrated into the body cues (using the traffic light system). For example, if they are asked to reach with the right hand, they are not to move their left foot. The right hand will be green, the left foot will be red, and the right foot and left hand will be yellow, see Figure 4. If during this task, the patient moved the a red limb (or in this case, the left foot) the limb will start flashing; until the movement is stopped. The feedback also works in the opposite direction, if a patient does not move the green limb (or in this case, the right hand) the limb will start flashing. This feedback is designed to stop the participant from walking back-and-forth across the space to reach the target and to prevent the participant from forgetting to reach/step for a target currently on the screen.

5 Conclusion

In this paper we described the design and development of a novel real-time Parkinson's Mediated Rehabilitation Environment that integrates sensory cueing techniques and the BIG protocol with visual and auditory feedback to create an engaging environment. Sensory cueing, informed by the BIG protocol, was used to encourage bigger and faster movement.

We believe the system has been designed to meet most of the user needs. In the coming months, we would like to perform a preliminary pilot study on able-bodied and PD participants. We plan to collect and analyze low-level movement data to further classify and analyze the movement dysfunction of PD.

The advantage of the current system is we can collect highly accurate movement data which we can used to analyze the effectiveness of the system. However, for future iterations and clinical implementation, this system should be simplified. We believe this setup can be simplified to use a single infrared camera instead of a Motion Capture System. The camera can be setup to capture the patient's silhouette, which can be used in place of the avatar in our system.

The system, as described, only provides rehabilitation in a single plane. To extend the system to a 3D environment, we have begun to explore the possibility of creating a floor based system; we believe two systems can be combined to create a 3D environment. This system can also be simplified to use a minimal number of cameras to capture movement from two different planes of movement, as an alternate solution.

Current literature has shown the high rate of success with visual and auditory cues in the lower extremities, particularly with respect to gait. However, it has not described the usefulness in the upper extremities. Therefore, future research should explore whether visual cues can increase the amplitude of movement in the upper and lower extremities in PD patients over the course of the physical therapy. We believe the usage of sensory cueing and the BIG protocol within a mediated environment has endless possibilities that we plan to explore.

References

1. Albani, G., Pignatti, R., Bertella, L., Priano, L., Semenza, C., Molinari, E., Riva, G., Mauro, A.: Common daily activities in the virtual environment: a preliminary study in parkinsonian patients. Neurol. Sci. 23(suppl. 2), S49–S50 (2002)
2. Azulay, J.-P., Mesure, S., Blin, O.: Influence of visual cues on gait in Parkinson's disease: contribution to attention or sensory dependence? J. Neurol. Sci. 248, 192–195 (2006)
3. Baker, K., Rochester, L., Nieuwboer, A.: The immediate effect of attentional, auditory, and a combined cue strategy on gait during single and dual tasks in Parkinson's disease. Arch. Phys. Med. Rehabil. 88, 1593–1600 (2007)
4. Camurri, A., Mazzarino, B., Volpe, G., Morasso, P., Priano, F., Re, C.: Application of multimedia techniques in the physical rehabilitation of Parkinson's patients. Journal of Visualization and Computer Animation 14, 269–278 (2003)
5. Chase, T.N.: The significance of continuous dopaminergic stimulation in the treatment of Parkinson's disease. Drugs 55(1), 1–9 (1998)
6. Dauer, W., Przedborski, S.: Parkinson's disease: mechanisms and models. Neuron. 39, 889–909 (2003)
7. Farley, B.G., Koshland, G.F.: Training BIG to move faster: the application of the speed-amplitude relation as a rehabilitation strategy for people with Parkinson's disease. Exp. Brain Res. 167, 462–467 (2005)
8. Farley, B.G., Fox, C.M., Ramig, L.O., McFarland, D.H.: Intensive Amplitude-specific Therapeutic Approaches for Parkinson's Disease: Toward a Neuroplasticity-principled Rehabilitation Model. Topics in Geriatric Rehabilitation 24, 99–114 (2008)
9. Ferrarin, M., Rabuffetti, M., Tettamanti, M., Pignatti, R., Mauro, A., Albani, G.: Effect of optical flow versus attentional strategy on gait in Parkinson's Disease: a study with a portable optical stimulating device. J. Neuroeng. Rehabil. 5, 3 (2008)
10. Freeman, J.S., Cody, F.W., Schady, W.: The influence of external timing cues upon the rhythm of voluntary movements in Parkinson's disease. J. Neurol. Neurosurg. Psychiatry 56, 1078–1084 (1993)
11. Harris, M.K., Shneyder, N., Borazanci, A., Korniychuk, E., Kelley, R.E., Minagar, A.: Movement disorders. Med. Clin. North Am. 93, 371–388 (2009)

12. Herman, T., Giladi, N., Hausdorff, J.: Treadmill training for the treatment of gait disturbances in people with Parkinson's disease: a mini-review. J. Neural. Transm. (Novomber 4, 2008)
13. Holden, M.: Cyberpsychology & behavior: the impact of the Internet, multimedia and virtual reality on behavior and society. Virtual environments for motor rehabilitation: review 8 (2005)
14. Huang, H., Wolf, S.L., He, J.: Recent developments in biofeedback for neuromotor rehabilitation. J. Neuroeng. Rehabil. 3, 11 (2006)
15. Jenner, P.: Molecular mechanisms of L-DOPA-induced dyskinesia. Nat. Rev. Neurosci. 9, 665–677 (2008)
16. Liebermann, D.G., Buchman, A.S., Franks, I.M.: Enhancement of motor rehabilitation through the use of information technologies. Clinical Biomechanics 21, 8–20 (2006)
17. McIntosh, G.C., Brown, S.H., Rice, R.R., Thaut, M.H.: Rhythmic auditory-motor facilitation of gait patterns in patients with Parkinson's disease. J. Neurol. Neurosurg. Psychiatry 62, 22–26 (1997)
18. Morris, M.E.: Movement disorders in people with Parkinson disease: a model for physical therapy. Phys. Ther. 80, 578–597 (2000)
19. Rubinstein, T.C., Giladi, N., Hausdorff, J.M.: The power of cueing to circumvent dopamine deficits: a review of physical therapy treatment of gait disturbances in Parkinson's disease. Mov. Disord. 17, 1148–1160 (2002)
20. Sidaway, B., Anderson, J., Danielson, G., Martin, L., Smith, G.: Effects of long-term gait training using visual cues in an individual with Parkinson disease. Phys. Ther. 86, 186–194 (2006)
21. Suteerawattananon, M., Morris, G.S., Etnyre, B.R., Jankovic, J., Protas, E.J.: Effects of visual and auditory cues on gait in individuals with Parkinson's disease. J. Neurol. Sci. 219, 63–69 (2004)
22. Trail, M., Protas, E., Lai, E.C.: Neurorehabilitation in Parkinson's Disease: An Evidence-Based Treatment Model, 1st edn. SLACK Incorporated, Thorofare (2008)
23. Sveistrup, H.: Motor rehabilitation using virtual reality. Journal of Neuro. Engineering and Rehabilitation 1, 10 (2004)
24. Winstein, C.J.: Knowledge of results and motor learning–implications for physical therapy. Phys. Ther. 71, 140–149 (1991)

Part II

VR for Culture and Entertainment

VClav 2.0 – System for Playing 3D Virtual Copy of a Historical Clavichord

Krzysztof Gardo and Ewa Lukasik

Institute of Computing Science, Poznan University of Technology,
60-965 Poznan, Poland
Krzysztof.Gardo@put.poznan.pl, Ewa.Lukasik@cs.put.poznan.pl

Abstract. VClav 2.0 system presented in the paper enables user to interact with a digital 3D reconstruction of the historical clavichord in a manner similar to lifelike using Virtual Reality gloves to "play" music. The real clavichord was constructed by the famous maker in 18th century, Johann Adolph Hass from Hamburg, and is at the exposition in the Museum of Musical Instruments in Poznan, Poland (department of the National Museum). This is a system powered by the NeoAxis game engine and equipped with 5DT Data Glove 14 and Polhemus Patriot tracker. It is an exemplary solution for museums to actively present musical instruments.

Keywords: cultural heritage, Virtual Reality, 3D modeling, clavichord, gesture driven HCI.

1 Introduction

Musical instruments represent very specific art objects in museum collections. Apart from being attractive as historical items of visual art, they are sources of sound. Their full appreciation is possible while listening to their music played live. Also an observation how the sound is produced is interesting for visitors. Musical instruments are interactive, but their interactivity may be hardly ever experienced in museums or because of their protection, or because their old mechanisms are entirely or partly damaged and the sound cannot be demonstrated. Most musical instruments at the exhibitions are mute. Replacing real objects by virtual ones – located in the natural scenery – constitutes an interesting solution opposing museum labels "do not touch". Advances in Virtual and Mixed Reality technologies give rise to new ventures devoted to the digitization and access to museum objects [8].

Mixed Reality (MR) is a term concerning visualization technique based on both real and virtual objects. It superimposes graphics, audio and other sense enhancements over a real-world environment in real-time. In general Mixed Reality requires three-dimensional interaction in which users' tasks are performed directly in a 3D spatial context [1]. It is possible to achieve using 2D input devices, but more desirable is using 3D technology. Typical 3D devices are 3D mouse, data gloves, pointer or gesture devices and other multiple DOF sensors or trackers available in the market or in the research labs. Unfortunately they come with a heavy price tag [3].

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 141–150, 2011.

Using Mixed Reality for presenting musical instruments is a challenge. Visual animated representations should be combined with sound and should enable real time interaction, so that the user has the impression of playing an authentic instrument. The research papers in the domain of digital reproduction of musical instruments concern rather the reproduction of their elements, as e.g. the two plates (top and back) of the soundboard of a historical violin [7] or an organization of virtual musical instruments exposition [6].

VClav 2.0 system presented in the paper is devoted to the precise visual reconstruction of the historical clavichord in the Virtual Reality with Virtual Reality data gloves to "play" music. The real clavichord was constructed by the famous 18th century maker, J.A. Hass, from Hamburg. The instrument was accidentally found, in a very bad condition, with many parts destroyed, and brought to the Museum of Musical Instruments in Poznan, Poland, for restoration. In order to make the audience acquainted with all interesting design solutions and decorative facets of the instrument and to show, how to play it, it was decided to perform a virtual reconstruction with new HCI interfaces.

A clavichord is a European stringed keyboard instrument known from the late Medieval to Classical eras [2]. It is still used in performances and recordings of renaissance and baroque music. It produces sound by striking brass or iron strings with small metal blades called tangents. Vibrations are transmitted through the bridge(s) to the soundboard. The action is simple, with the keys being levers with a small brass tangent at the far end.

The paper is structured as follows: Section 2 recapitulates the works on the 3D model of Hass clavichord undertaken within VClav and VClav 2.0 projects, Section 3 presents Virtual Reality hardware and software used in VClav 2.0, Section 4 desribes the system for playing a 3D virtual copy of Hass clavichord and Section 5 concludes the paper.

2 From VClav to VClav 2.0

The main objectives of the VClav 2.0 project were to create the faithful 3D model of the J.A. Hass clavichord with sound generating mechanism in operation and to introduce a realistic interaction between a user and an on-screen virtual instrument. The project is an extended version of VClav system (described in [5]), written in Java, where the 3D model of the clavichord was embedded. A great effort has been put to the photorealistic 3D modeling of the clavichord and to activate its movable parts. The process of virtual reconstruction of the instrument was thoroughly planned and followed part by part in the manner similar to the real one. The 3D model of the clavichord was based on the blueprint scheme, photographic documentation and on the actual instrument inspection. It was built part-by-part by the University team and covered with faithful textures, including valuable paintings on the soundboard. The 3D digital model was embedded in a standalone Java application written from scratch by the team. It enables a user to manipulate instrument's parts and uncover its certain layers. The interaction with the digital model is possible using a 3D mouse - a space navigator, together with a traditional mouse. Although the functionality of VClav was simplified, several of its stages, mainly modeling in Autodesk 3ds Max constituted the base for VClav 2.0 tasks.

The whole reconstruction process included the following steps:

- analysis of the construction of the instrument and planning the digital reconstruction steps,
- modeling the exact copy of the parts of the instrument in 3D software environment,
- digitally reproducing historical textures,
- reconstructing missing parts and digitally “repairing” destroyed parts,
- choosing the VR tools necessary to provide the assumed functionality and interface,
- creating the digital environment to manipulate and to play the digital clavichord,
- creating or adapting the physics engine, i.e. computer software that provides an approximate simulation of certain physical systems, such as rigid and soft object dynamics in real-time and equipping objects with basic physical features.

VClav 2.0 significantly extends the functionality of the first version. All operations are executed in a realistic manner – a player, wearing a data glove, has his/her hand virtually reproduced on the screen and plays the virtual instrument, as a real one.

The extended functionality and real-time operation demanded using an advanced software tools that would simulate the performance, actions and physical phenomena while operating various parts of the virtual instrument. An existing systems have been considered, including physics and game engines [4].

A physics engine is computer software that provides an approximate simulation of certain simple physical systems, such as rigid body dynamics (including collision detection), soft body dynamics, and fluid dynamics, of use in the domains of computer graphics, video games and film. Their main uses are in video games (typically as middleware), in which case the simulations are in real-time.

A physics engine often constitutes a core part of a more advanced system – a game engine. A game engine is a software system designed for the creation and development of video games. The core functionality typically provided by a game engine includes a rendering engine (“renderer”) for 2D or 3D graphics, a physics engine, i.e. a module simulating physical phenomena or collision detection (and collision response), a control module receiving signals from input devices (a keyboard, a mouse, a joystick, etc.) and a sound controlling module. The above-mentioned features of the game engine may be applied using the Application Programming Interface provided by the engine, however some of them are so advanced that they make available a complete set of tools for producing games via the software development kit (SDK) including all the indispensable libraries for scripting, animation, artificial intelligence, networking, streaming, memory management, threading, localization support, and a scene graph. Some engines are equipped with converters that allow the programmer to rewrite data from external programs to make them compatible with the modules of the engine.

The following requirements have been imposed on the game engine to be used within the VClav 2.0 project:

- programming language and IDE: C #, MS Visual Studio 2008,
- independence from the graphics driver: OpenGL and DirectX 9,
- built-in physics and sound system.

Searching for the engine was carried out using DevMaster.Net website, which contains a database of game engines with advanced search engine [10]. After a thorough analysis of the available game development software, the NeoAxis engine in its non-commercial version (described in the Section 3) was chosen in the form of the middleware software. The engine was responsible for displaying three-dimensional graphics, playing sounds, and simulating the software environment for the realization of physics laws provided by a physics simulation engine NVIDIA PhysX.

3D models of the clavichord parts, the scenes, and the virtual hand were exported as models compatible with the NeoAxis engine where motion was added to various parts of the clavichord and a hand.

VClav 2.0 system operates in two modes:

- a 3D clavichord model manipulation and
- playing the instrument.

The manipulation mode allows the user to control moving parts of the instrument (e.g. opening and closing the covers) and using his/her own hand represented as the model on the screen. In the game mode a player wearing the glove can "hit" the keys of the clavichord by the virtual hand fingers, activate the key levers with tangents and generate the sound. The WAVE samples of a real clavichord [9] sound are used to play music.

3 Virtual Reality Hardware and Software

The project VClav 2.0 uses two Virtual Reality input devices: 5DT Data Glove Ultra 14 and Polhemus Patriot tracker. As a display 3D options have been considered, but finally a full HD 2D monitor was used as most available at the time of finalizing the project. Fig. 1 presents the development environment for VClav 2.0.

The target operating system of the project is Microsoft Windows XP/Vista/7 installed along with Microsoft. NET Framework 3.5 [11]. The project was written in C# and prepared in Microsoft Visual Studio 2008 integrated development environment. Its implementation was based on the NeoAxis Engine(Non-Commercial SDK 0.8.4.2 [14]) and libraries and controllers of the VR tools used in the system. For designing 3D elements Autodesk 3ds Max 2010 was used.

3.1 5DT Data Glove Ultra 14

5DT Data Glove 14 Ultra is a 14-sensor data glove, equipped in two sensors per finger, that measure finger flexure as well as the abduction between fingers [17]. There are USB and wireless version of gloves. Sensors output value is in the range 0-4095. The glove was tested by static and dynamic tests. Static tests were performed on motionless hand. The stability of sensors output was different for each of them, but acceptable for assumed fingers motions. Dynamic tests were executed for flexures of individual fingers and have also shown certain amount of variability (acceptable).

3.2 Polhemus Patriot Tracker

Polhemus Patriot tracker [16] utilizes AC tracking technology. The source and sensor contain electromagnetic coils enclosed in plastic shells. The source emits magnetic fields, which are detected by the sensor. The sensor's position and orientation are measured as it is moved. It provides dynamic, real-time measurements of position (X, Y and Z Cartesian coordinates). Update rate is 60Hz per sensor., latency - less than 18.5ms. The tracker was tested while moving along the three coordinates and showed the greatest error along Z-coordinate. However playing the keyboard instrument hardly involves movement along Z-coordinate, therefore the results were considered as acceptable.

3.3 The Display

The digital reconstruction of the clavichord was devoted for the exploitation in the Museum of Musical Instruments, so it had to offer a full HD image quality to present the details of construction and ornamentation. Although a 3D display has been considered as a presentation device [1], including a the Head Mounted Displays, it was finally decided to use traditional full HD two-dimensional 40” monitor, as best suited to the existing setting in the Museum and most available at the time of finalizing the project.

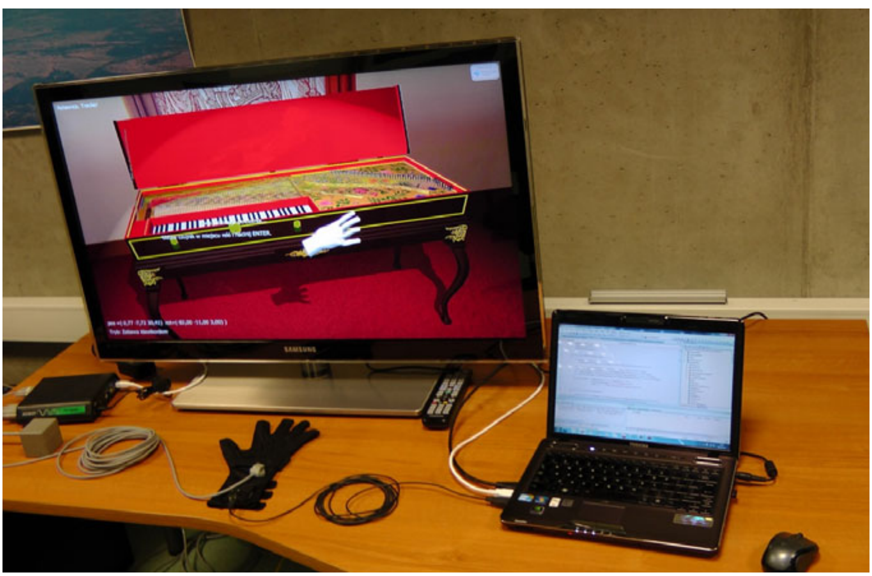


Fig. 1. VClav 2.0 development environment: Samsung LED 40’ display, laptop Toshiba U500 Series, Patriot Polhemus Tracker and 5DT Data Glove 14 Ultra

3.4 NeoAxis Engine

NeoAxis Engine is a complete integrated development environment for creating interactive 3D graphics including 3D virtual worlds, AAA games, and realistic simulations [14]. The system comprises a real-time 3D engine and a suite of full featured tools. It uses OpenGL graphics and MS. Direct3D drivers to display graphics. It is integrated with PhysX technology and ODE Physics, which allow the designer to simulate physical phenomena for the displayed scenes. To play the sound DirectX Sound System or OpenAL can be used. Non commercial version with limited functionality has been used in the project. From a numerous tools of the engine, the following have been used:

- *Resource Editor* - a tool for viewing and editing partial resources used in the project, i.e.: 3D objects, states of system physics, sound or special effects.
- *Map Editor* - the integrated object placement toolkit for the easy building of game scenes.
- *NeoAxis Engine Exporter* - model exporters, supporting popular 3D packages, including Autodesk 3ds Max.

NeoAxis Engine Exporter is a tool to convert models from Autodesk 3ds Max to the form compatible with OGRE standard used within the NeoAxis Engine. This tool is able to create a mesh with the material and then apply it to the model. Exporter recognizes the backbone of the model together with the animation, which can be used in the project.

4 System for Playing a 3D Virtual Copy of a Historical Clavichord

The work related to the project implementation was carried out in several consecutive stages and in a particular order. The first stage consisted of modifying three-dimensional VClav models of clavichord, scenes, and virtual hand for the new environment of NeoAxis engine. Second and third stages were associated with the creation of physics model of the clavichord and the virtual hand. In the last stage of the project the application code has been written.

4.1 Modifying the 3ds Max Models to the NeoAxis Engine Format

Model of the Clavichord and a Scene. Every functional element of the virtual clavichord had to be the individual 3D object (drawer, key, lid). A great deal of modifications was performed to the VClav model elements, as some of them were composed of several parts, e.g. the key consisted of five separate shapes: white and black part, and three wooden components and a tangent. All of them had to be merged into one object.

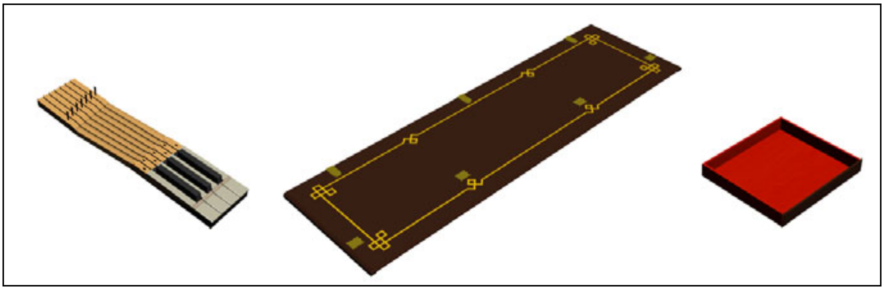


Fig. 2. Dynamic objects in VClav 2.0: keys, the lid, the drawer

Fig. 3. Static objects in VClav 2.0: the casing and the rack

Before exporting the objects to NeoAxis engine, they were also equipped with the additional features: texture format, animation switching on/off, switching on shadows, gravitation and collision between the objects. At this stage the names of objects were adapted to the new environment. Particularly important was assigning numbers to the keys, because depending on their number, a sound of a specific pitch is played. These elements are only presentation parts used by the engine. Apart from them every object has its shape, which is taken into account during detection of collision or simulation of the physical environment. Fig. 2 and Fig. 3 present examples of dynamic (movable) and static objects in VClav 2.0.

Model of the Virtual Hand. The 3D model of a hand comes from the free TurboSquid service [15]. Its modification consisted in its equipment with the skeleton system of fingers. In the 3ds Max it may be realized using the tool "Bones" (Create->System->Standard). It is necessary to point out the beginning and the end of the bone. Program merges bones situated close one to the other. Some difficulty was related to adjusting bones system to the shape of the hand, which in this case is in an unusual arrangement of fingers – while playing the keyboard. It had also to be rescaled to match the size of the clavichord.

4.2 Definition of Physical Mechanisms of the Clavichord and a Hand

In NeoAxis Engine there exists systematic and hierarchic structure of resources. The highest in the hierarchy of resources is a Map. First the designer has to create a map, which is the starting point to display a scene. The Map consists of units called Entities. The Entity can be shaped as a mesh and imposed on its Material. Other Entities are: light, camera, particle effects or sound. The important Entity is the type instance (Type). The Type combines a mesh covered with Material, Sound and Physics Model.

To be able to create a virtual environment in which physical dependencies between objects, similar to the real ones exist, the set of rules has to be defined. These rules form the Physics Model. The Physics Model is composed of Bodies, joints between the bodies (Joints), and the engines and motors (Motors). Each body defines a shape (Shape) and describes its physical properties such as static friction, dynamic friction, center of gravity, the material (based on the calculated mass). The shape of the body consists of several basic shapes (Shape). Basic shape may be one of the structures such as rectangular, capsule, sphere, and mesh (Mesh Triangle) - which in a non-commercial version of the engine have been blocked. It is an important limitation, as it was impossible to construct clavichord physics using the mesh of the instrument model. Instead the model of the instrument (casing, keys, cover, drawer and others) was constructed in the 3d studio Max using exclusively cuboids. Some additional objects contained information concerning rotation axis of keys and places, in which lid hinges were placed. The NeoAxis engine is not equipped in the tool that would enable the export of physics models from the 3ds Max, that is why a special script, MaxScript, describing cuboid data in the text file, was created. Then the "raw" data representing shapes were re-writen to the file with the notation concordant with the description of the model in the NeoAxis engine.

NeoAxis engine is equipped in a system simulating physics phenomena. Initially the hand and fingers models constituted a complex system of bones joints, and motion engines of ServoMotor type. However due to some unacceptably long reaction times the function of bending had to be created from scratch. The following operations were created: moving and rotating a hand and fingers as well as bending fingers. These operations demanded from the system co-sharing additional information describing the intermediate states of the elements of the object.

4.3 Implementation of the Mechanism Controlling the Glove and the Tracker

The control mechanism of input devices converts data coming from the tracker and from the glove in such a way as to allow visual mapping of the location, slope and a state of the user hand. Received data are subject to transformation according to the calibration data.

All glove sensor data are standardized in such a way that the value equal to 0 responds for the natural flexion of the finger. Maximum flexion of the finger responds to the sensor value equal to 100%.

4.4 VClav2.0 in Use

The main interface to the system is composed of the Data Glove and the Tracker. The keyboard is used for some basic commands. Moving a virtual hand is accomplished by moving the user's hand, on which the tracker sensor is placed. Bending virtual fingers is realized by bending the fingers of the user hand wearing 5DT Data Glove.

Navigating in the virtual world is via a virtual white hand in two modes:

– a 3D clavichord model manipulation,
– playing the instrument.

In the first mode the virtual hand opens the lids of the instrument. To enter into the playing mode the user has to press [SPACE] on the computer keyboard. The camera is moved over the keys and fingers may push the keys to play the instrument. (Fig. 4). Consecutive pressing [SPACE] will restore the view of the whole instrument.

Pressing the key by the finger of the virtual hand in playing mode causes its rear part to rise so the tangent hits the string. In real instrument the string is set into vibration and the sound is produced. In the virtual version the sound is played from samples when the rear part of the virtual key reaches the appropriate height.

Fig. 4. The scene view in a *play* mode

5 Conclusions

The goal of the project presented in this paper was to popularize the valuable 18th century clavichord of Johann Adolph Hass, that will never be playing in a reality due to its severe damage. In the search of the method to virtually present the clavichord, its construction and the way to generate sound, the idea was realized, to make the virtual 3D copy of the instrument and to equip the user with the virtual data glove to enable him/her to virtually play the instrument. VClav 2.0 system has been built powered by the NeoAxis game engine and equipped with 5DT Data Glove 14 and Polhemus Patriot tracker. Since the NeoAxis engine was applied in its non-commercial version, several functions were programmed from scratch. A series of user tests were accomplished including public active demonstrations. VClav 2.0 found appreciation of both – users and the staff of the Museum of Musical Instruments. However the data glove appeared not enough sensible and precise to make the user feel comfortable while touching keys. The results seem promising for further development of the project which will go towards more robust solution.

Acknowledgements. The authors would like to thank Dr. Mikolaj Sobczak, leader of the Mobile Systems Research Labs, Poznan University of Technology [12] for providing the VR equipment without which the VClav 2.0 project could not have been accomplished.

Special thanks go to Mirosław Baran, Piotr Cieslak, Patryk Frankowski, Maciej Szarafinski and Janusz Jaskulski from the Museum of Musical Instruments, National Museum, Poznan, Poland [13] for their assistance during all stages of the project.

References

1. Bowman, D.A., Chen, J., Wingrave, C.A., Lucas, J., Ray, A., Polys, N.F., Li, Q., Haciahmetoglu, Y., Kim, J.S., Kim, S., Boehringer, S., Ni, T.: New directions in 3D user interfaces. The International Journal of Virtual Reality 5(2), 3–14 (2006)
2. Brauchli, B.: The Clavichord. Cambridge University Press, Cambridge (2005)
3. Chow, Y.-W.: Low-Cost Multiple Degrees-of-Freedom Optical Tracking for 3D Interaction in Head-Mounted Display Virtual Reality. International Journal of Recent Trends in Engineering 1(1) (2009)
4. Craig, A., Sherman, W.R., Will, J.D.: Developing Virtual Reality Applications, Foundation of Effective Design. Morgan Kaufmann Publishers, San Francisco (2009)
5. Gardo. K., Nowaczyk K., Przewozna, D., Lukasik E.: Digital Model of Hass Clavichord A.D. 1754, accepted for DMACH (2011)
6. Pavlidis, G., Tsiafakis, D., Provopoulos, G., Chatzopoulos, S., Arnaoutoglou, F., Chamzas, C.: MOMI: A dynamic and internet-based 3D virtual museum of musical instruments. In: Third International Conference of Museology, Mytilene, Greece (2006)
7. Pinto, L., Roncella, R., Forlani, G.: Photogrammetric survey of ancients musical instruments. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS Congress, vol. XXXVII, Part B5, pp. 308–313 (2008)
8. Walczak, K., Cellary, W., White, M.: Virtual museum exhibitions. IEEE Computer 39(3), 93–95 (2006)
9. Clavichord samples, `http://www.hauptwerk.cz/klavichord/index.htm`
10. DevMaster.Net - 3DEngines Database, `http://www.devmaster.net/engines`
11. Microsoft .Net 3.5 Runtime Environment, `http://msdn.microsoft.com/en-us/netframework/cc378097.asp`
12. Mobile Systems Research Labs, `http://mlabs.pl/`
13. Museum of Musical Instruments (O/MNP), `http://www.mnp.art.pl/oddzialy/mim/index.php`
14. NeoAxis - Documentation, `http://www.neoaxisgroup.com/wiki/Documentation`
15. TurboSquid, repository of 3D objects, `http://www.turbosquid.com`
16. Polhemus, `http://www.polhemus.com`
17. 5DT Data Glove, `http://www.5dt.com/products/pdataglovemri.html`

A System for Creating the Content for a Multi-sensory Theater

Koichi Hirota[1], Seichiro Ebisawa[1], Tomohiro Amemiya[2], and Yasushi Ikei[3]

[1] Graduate School of Frontier Sciences, University of Tokyo,
5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8563, Japan
[2] NTT Communication Science Laboratories, NTT Corporation,
3-1 Morinosato Wakamiya Atsugi-shi, Kanagawa, 243-0198, Japan
[3] Faculty of System Design, Tokyo Metropolitan University,
6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan
hirota@media.k.u-tokyo.ac.jp, amemiya.tomohiro@lab.ntt.co.jp,
ikei@computer.org

Abstract. This paper reports on the current progress in a project to develop a multi-sensory theater. The project is focused not only on the development of hardware devices for multi-sensory presentations but also on an investigation into the framework and method of expression for creating the content. Olfactory, wind, and pneumatic devices that present the sensation of odor, wind and gusts, respectively, were developed and integrated into an audio-visual theater environment. All the devices, including the video device, are controlled through a MIDI interface. Also, a framework for creating the multi-sensory content by programming the sequence of device operations was proposed and implemented.

Keywords: multi-sensory theater, odor, sensation of wind, multi-sensory content.

1 Introduction

Television networks have effectively served as the framework for distributing audio-visual content. Penetration by the Internet and the spread of personal computers are expected to enable the delivery of multi-sensory content (e.g. including force, tactile, and olfactory sensations), in addition to audio-visual content, because more complicated processing of multi-sensory information is possible at the end of a communication line. With this background, the authors have been carrying out a project aimed at developing a multi-sensory theater and the technology needed to exploit it.

Multi-sensory theater, in our definition, is an expansion of home theater, and is an environment that provides tactile, force, olfactory, and vestibular sensations (see Figure 1). The theater is expected to serve as an environment for viewing passive content, experiencing interactive content, and communication.

The project is also investigating a framework for creating multi-sensory content and technology to support the process of creation. Just as video editing software has

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 151–157, 2011.

simplified the creation of video content, a standard framework to deal with multi-sensory materials is expected to stimulate the creation and distribution of multi-sensory content. It should be noted that the multi-sensory theater in this case serves as an environment for monitoring the creative content. For this purpose, the theater must be simple and easy to use, and a systematic approach towards the effective expression and integration of multi-sensory information needs to be found.

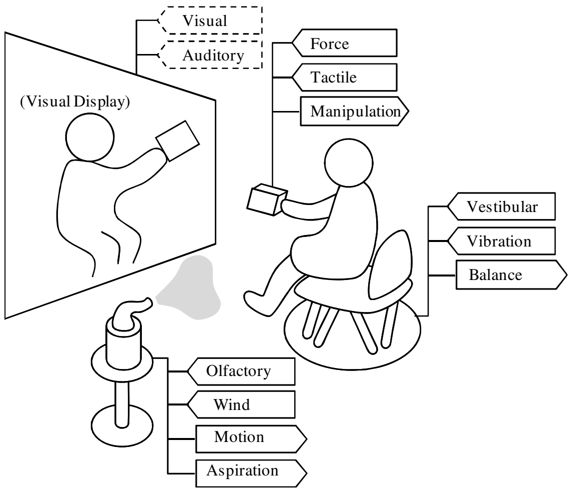


Fig. 1. Complete image of multi-sensory theater in our project. The theater is intended to be used not only for presentation of passive content but also for interactive content and communication. This is the reason why it is planned to integrate some input devices into the theater.

2 Related Research

Pioneering work on multi-sensory theater was done by Heilig[1] who developed a system known as the Sensorama Machine, which provides the complex experience of riding a motorcycle by presenting wind, odor, and vibration in addition to the visual and audio sensations. It is expected that the concept of multi-sensory presentation will be expanded for general-purpose use by using modern computers and sensory device technologies.

Recently, various attractions in amusement parks have taken advantage of multi-sensory presentation in addition to audio-visual sensations [2]. Especially, vestibular sensations have been effectively used for ride-type attractions to provide the sensation of riding a vehicle. Although not reported in this paper, our project is also planning to integrate a device for vestibular sensation into the theater.

The presentation of olfactory sensations is considered to be still in the research stage rather than in practical use. Although, ultimately the presentation of odor is likely to be through the real-time synthesis of different chemical substances, most current research is focused on just mixing odorants[3-7]. Even in the mixing approach, if a sufficient variety of odorants is prepared, it becomes viable for practical applications. For our project we decided to use a device based on the mixing approach[8].

From the view point of relatedness to the visual scene, the sensation of wind is expected to have the effect of adding reality. This is considered to be the reason why the sensation of wind was integrated in the Sensorama Machine. The precise reproduction of air flow is considered to be an inverse problem of fluid dynamics that seeks the boundary conditions of a space when the flow inside the space is given. Most research on wind devices in human-computer interactions, however, are aimed at presenting an approximation to the air flow, such as the direction and intensity of the wind[9-11]. Our research also employs this approximation approach.

The goal of multi-sensory interfaces is frequently understood to be an accurate reproduction of real sensations using high-fidelity display devices. This approach is considered to be focused on reality at the physical level in Figure 2. However, the actual goal of the interface is usually transmission of higher level information. Also, since the progress in developing devices is not very fast, the attainment of a high sense of reality at the physical level is considered to be improbable in the near future. Moreover, there is some content that has less of a relationship with reality, such as surreal or fictional content. Hence our project is aimed at investigating presentation and expression at the levels of sensation and scene in Figure 2.

This paper reports on the progress of the project including the implementation of a prototype theater environment and the development of a framework for editing multi-sensory content.

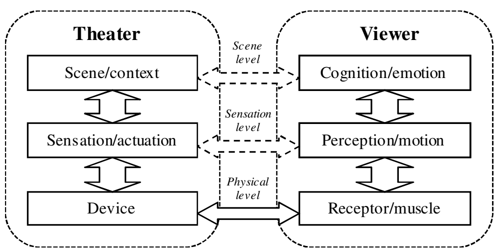


Fig. 2. Schematic of information transmission between the theater and the viewer. The contact point between the theater and the viewer is at the physical level of the hierarchy. However, the purpose of content creation is usually the transmission of information at the sensation or scene level.

3 Multi-sensory Theater and Content Editing System

Our prototype theater environment is currently composed of audio-visual, wind, olfactory, and pneumatic devices (Figure 3). Since there are no commercially available standard devices that present sensations other than auditory and visual sensations, it is necessary to develop new devices and to design a system into which the devices can be integrated.

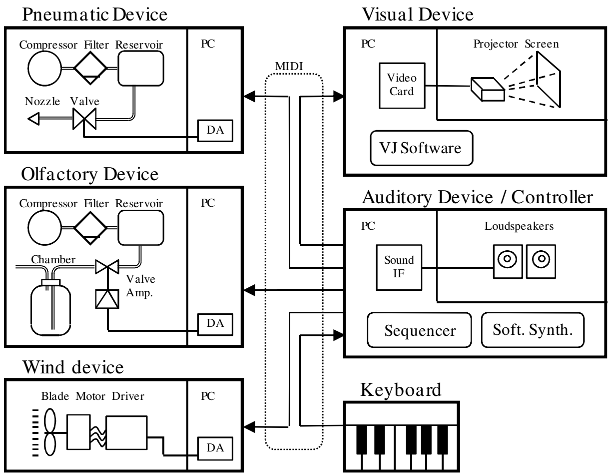


Fig. 3. Arrangement of devices in the prototype theater. Each device has been implemented as a MIDI device and is controlled by a sequencer.

3.1 Devices and Theater Environment

Olfactory devices present odors by emitting an odorant vaporized in a chamber by injecting compressed air into the chamber. By controlling the volume of injected air, the intensity of the odor in the theater environment can be changed. Although the device has 16 output channels, in our preliminary experiment only 4 channels, which present rose, peppermint, hay, and grapefruit aromas, were used.

The wind device is a type of electric fan whose rotational speed can be controlled by a computer; the maximum velocity of the wind is approximately 3 m/s at 3000 rpm. The current theater uses 4 channels out of a maximum of 8. The changes in wind intensity vary depending on the generation mechanism. In the current implementation, control of the intensity (i.e., rotational speed) is based on a trapezoidal velocity pattern whose acceleration and deceleration times are given as control parameters. Changing the acceleration and deceleration times empirically gives quite different impressions of the nature of the wind.

The pneumatic device presents wind by ejecting compressed air from nozzles. It controls the intensity of the air flow by controlling the pressure of the compressed air. This device is capable of presenting a forceful wind and gusts although a quantitative assessment of the wind velocity has not yet been done. The device has 16 output channels whose nozzles are, in the current installation, arranged in an array and directed toward the user's face.

The audio-visual system consists of a projector and a screen (80 inch diagonal) and a set of stereo loudspeakers. The theater environment has been constructed by integrating all the devices described above into the audio-visual environment. An approximate layout of the environment is shown in Figure 4. The distance from the screen to the user is relatively long compared with usual theaters because some space has been left to facilitate maintenance. In a future revision, the distance will be reduced to provide a wider field of view to the user.

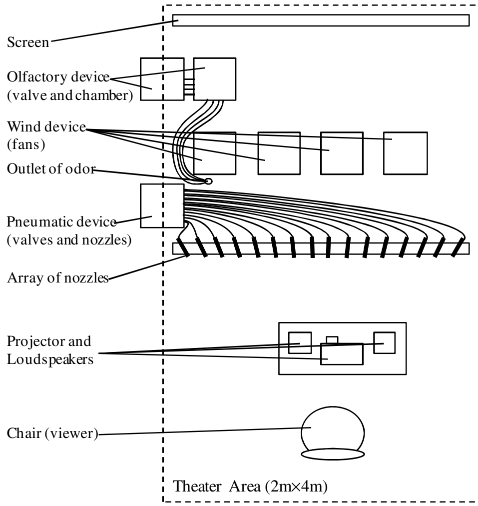


Fig. 4. Layout of the prototype theater environment. Olfactory, wind, and pneumatic devices are arranged in front of the viewer.

3.2 Content Editing Framework

There is no standard format or framework for transmitting and distributing multi-sensory content. Even in the experimental creation of multi-sensory content in our project, a framework for describing and playing back content is required. Installation of the hardware devices only is not sufficient for it to serve as a theater.

In our study, a MIDI-based control system was implemented, where all the devices have a MIDI interface, and commands and parameters are transmitted to those devices by a MIDI sequencer. The sequencer also provides the interface for editing the sequence program including real-time recording using MIDI instruments (e.g. a musical keyboard), inputting and adjusting on a timeline, and balance control between multiple tracks.

As was mentioned at the beginning of this paper, the multi-sensory theater is an environment for editing and monitoring content as well as an environment for viewing. An advantage of the implementation described above is that, because of the nature of the MIDI system, the results of all editing operations can be immediately reflected back from the devices to the editor.

4 Editing Process

In a preliminary experiment towards a case study, some simple multi-sensory sequences were created. This is described in the following steps. An example of the resulting sequence in piano roll representation is shown in Figure 5.

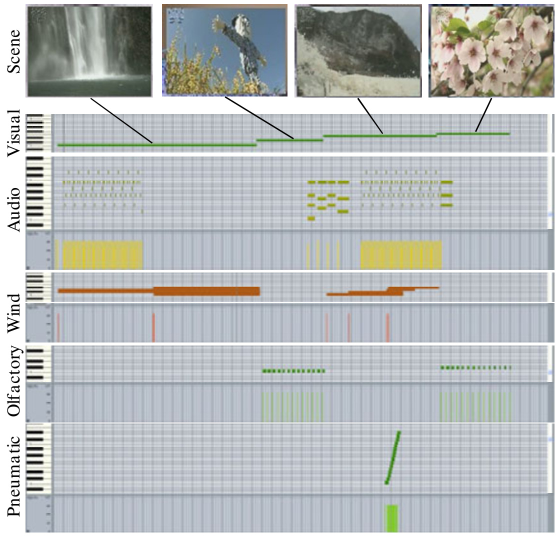


Fig. 5. Sequence data from a simple program. The program consists of four scenes. The sensations of odor, wind, and gusts have been added depending on what is depicted in each scene.

1. Audio-Visual Tracks. First, audio material (i.e. MIDI music data in this example) was loaded to the sequencer software to create audio tracks; in the figure only one track is shown although the entire music data was composed of 6 tracks; the bar chart below the piano roll shows information on velocity (i.e. intensity of sound). Also, video material (i.e. short movies) was loaded into the VJ software of the video device, which enabled playback of those video images and producing effects among them in real-time according to the control from the sequencer software. Then the video sequence was determined by trial and error by punching in in real time to the music. As shown in the video track, four scenes were produced in sequence.
2. Olfactory and Wind Tracks. The odor and wind sequence was then embedded into the audio-visual scenes. Also in this process, the function of recording events in real time was used effectively. In scenes 2 and 4 in Figure 5, the odors of hay and rose, respectively, were emitted to add atmosphere to those scenes. Since the odor must last through the scene, the odorant was emitted intermittently during the scene. Wind was presented to add reality to the images in scenes 1, 2, and 4. Gusts generated by a pneumatic device were used only with scene 1 to simulate the sensation of splashes from waves; temporally changing the channels meant that the direction of the wind that the user felt changed quickly from one side to another.
3. Fine Adjustment. Finally, adjustment of all the input data on timing and intensity was performed. Mistakes in real-time recording were corrected and subtle variations in intensity were fixed on the time-line (i.e. the piano roll). Although not used in this example, it is also possible to copy and paste a part of sequence to another part to reuse common expressions.

5 Conclusion

This paper reports on our study on the implementation of a multi-sensory theater and a framework for creating multi-sensory content. Through experimental editing of short sequence programs, it was shown that the proposed framework can provide a minimum set of functions required for content creation. To achieve the goal of the project, some devices still need to be integrated into the theater. Also, improvement and evaluation of the framework for content creation is required to make the framework available for practical use.

Acknowledgement. This research was carried out as part of a project that is supported by the National Institute of Information and Communication Technology (NICT).

References

1. Heilig, M.L.: The Father of Virtual Reality, http://www.mortonheilig.com/
2. Walt Disney World, http://disneyworld.disney.go.com/parks/
3. Hiramatsu, T., Nakahira, Y., Nakamoto, T., Moriizumi, T.: Study of odor recorder using active odor sensing system; IEICE technical report, CPM2000-32, 17-22 (2000)
4. Kadowaki, A., Sato, J., Bannai, Y., Okada, K.: Measurement of Olfactory Responses to Odor Stimulation Switched in Single Breath. J. Japan Association on Odor Environment 40(1), 1–10 (2009) (in Japanese)
5. Yamada, T., Yokoyama, S., Tanikawa, T., Hirota, K., Hirose, M.: Wearable Olfactory Display: Using Odor in Outdoor Environment. In: Proc. IEEE Virtual Reality 2006, pp. 199–206 (2006)
6. Yanagida, Y., Kawato, S., Noma, H., Tomono, A., Tesutani, N.: Projection based olfactory display with nose tracking. In: Proc. IEEE Virtual Reality, pp. 43–50 (2004)
7. NiCT press release: Development of multi-sensory interaction system integrating four sensations, http://www2.nict.go.jp/pub/whatsnew/press/h21/091104/091104.html
8. Hirota, K., Ikei, Y., Amemiya, T.: Olfactory Display for Multi-sensory Theater. In: Proceedings of Asiagraph 2010, pp. 24–28 (2010)
9. Ogi, T., Hirose, M.: Usage of multisensory information in scientific data sensualization. Journal of Multimedia Systems 5(2), 86–92 (1997)
10. Sawada, E., et al.: BYU-BYU-View: a wind communication interface. In: SIGGRAPH 2007 ACM SIGGRAPH 2007 Emerging Technologies (2007)
11. Kosaka, T.: WindStage (WindDisplay & WindCamera). The Journal of the Society for Art and Science 8(2), 57–65 (2009) (in Japanese)

Wearable Display System for Handing Down Intangible Cultural Heritage

Atsushi Hiyama[1], Yusuke Doyama[1], Mariko Miyashita[2], Eikan Ebuchi[3], Masazumi Seki[3], and Michitaka Hirose[1]

[1] Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan
[2] Tamagawa University, 6-1-1, Tamagawagakuen, Machida-shi, Tokyo, Japan
[3] Kochi Prefectural Paper Industry Technology Center, 287-4, Hakawa, Ino-cho, Agawa-gun, Kochi, Japan
{atsushi,ydoya,hirose}@cyber.t.u-tokyo.ac.jp

Abstract. In recent years, most of traditional craftsmanship is declining because of aging skilled craftspeople and fewer successors. Therefore, methods for digital archiving of such traditional craftsmanship are needed. We have constructed a wearable skill handing down system focused on first-person visual and audio information and biological information of a craftsman. We used instrumental information associated with the usage of the tools for evaluating the effect of proposed wearable display system of intangible cultural heritage. In this paper, we show the result of archiving and training on the skills of Kamisuki, Japanese traditional papermaking.

Keywords: Intangible cultural heritage, Skill transfer, Tacit Knowledge, Wearable computer.

1 Introduction

These days most of traditional craftsmanship has declined to the brink of extinction, mainly due to shrinking number of successors and aging experts. Even though there are urgent needs for preservation of declining traditional craftsmanship, it takes quite a long time to transfer skills from experts to learners with conventional methods. Traditionally, these skill-transfers have been carried out by word-of-mouth between a master and apprentices, relying heavily on the existence of the master as a teacher. Within this teaching framework, only limited number of apprentices can learn the craftsmanship.

Therefore, new methods for archiving and transferring of skills that do not entirely depend on the existence of experts are required. Although it is possible to record these skills in a form of videos or textbooks with pictures, those visual information do not contain enough information to master these skills. In order to achieve a better

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 158–166, 2011.

understanding of the skills, it is necessary to have invisible information that is not observable from outside.

So far, little research has been done on digital preservation of traditional craftwork skills. There is plenty of research on digital preservation of static cultural heritage such as paintings, craftworks and buildings [1, 2, 3]. Some researches focus on the artistic movements such as dance [4]. However, focus of these studies is on the existing art objects or artistic movements themselves, not on the artisanship to create those art objects. Saga et al. developed a unique system to transmit calligraphic skills using haptic feedback [5]. Our approach takes a similar attitude toward preservation and transmission of skills, but in a more integrated way with various sensors, taking deeper look at the tacit knowledge of experts numerically.

We have created an integrated framework for the digital preservation of artisanship [6, 7]. In this paper, we explain our framework for digital archiving of craftwork skills, and show a verification experiments in which we applied the system to a Japanese traditional craftworks, Kamisuki.

2 Kamisuki, Japanese Traditional Paper-Making

As the research object, we chose Kamisuki, Japanese traditional papermaking. We conducted pilot survey for preservation of this craftsmanship.

Traditional papermaking has more than 1,300 years of history. After the industry reached its peak with 68,562 production sites in the beginning of the 20th century, it followed a course of decline and the number of production sites decreased to 392 [8]. Even though traditional Japanese paper gains attention again in recent years, acceptance of new apprentices is difficult for most of the production sites which are family-owned small factories. Therefore, new framework for digital archiving and transferring of skills are needed.

The whole process of papermaking starts with the preparation of materials. Raw materials such as kozo, gampi and mitsumata go through boiling, bleaching and cleaning. Then those materials are crashed into a collection of fiber and put to water with *Neri*, viscosity improver made of grass roots. The papermaking part follows. This part consists of three main movements: scooping, swinging and draining. Figure 1 shows those three main movements.

Fig. 1. Three main movements in Japanese papermaking; scooping, swinging and draining (from the left)

We chose the papermaking part as a research object, because it determines the quality of the final product as well as it is a combination of dynamic movement and subtle control of the body. In order to capture such dynamic movement and subtlety at the same time, our approach seemed suitable.

3 Methodology

We focused on the development of the system to capture integrated information of the working environment, including visual information of the workspace, biological information of a craftsman, and instrumental information on the physical quantities of tools. The biological information includes the artisan's myoelectric signals, breathing and the eye-gaze during creation. On the other hand, the instrumental information of the tools is their physical quantities such as their position and load on them. It is important to record such invisible information numerically, because even experts are not consciously aware of how they moves their body and how they handle their tools, thus they do not have precise words to describe their movement.

The aim of this framework is summarized into three parts. First of all, we want to archive the knowledge as a whole. We want to capture as much information as possible, so that we can hand down enough information on the craftworks for future generations. Secondly, we want to find cause-and-effect relationships among those data. Those relationships among data considered as tacit knowledge that has never been clearly analyzed. Finally, we want to develop new skill-transfer methods for traditional craftworks with the help of digital technology.

We constructed HD camera (Canon XH-G1s). Camera captures the movement of the master and his way of handling tools through several viewpoints.

As a key tool for capturing biological data, myoelectric potential was measured with Nihon Koden WEB-100. Sensors on the artisan send information remotely, so that they do not bother craftspeople with cables.

Instrumental information is captured through various ways depending on the types of tools the craftwork adopts. Physical quantities are mainly measured, such as position, acceleration and distribution of load.

4 Measuring Kamisuki Skills

We conducted pilot servey with two subjects. Subject A has about thirty years of experience in this field. Subject B has about ten years of experience. We captured the data while both subjects make 10 pieces of paper. Sensors were set at the position shown in Figure 2.

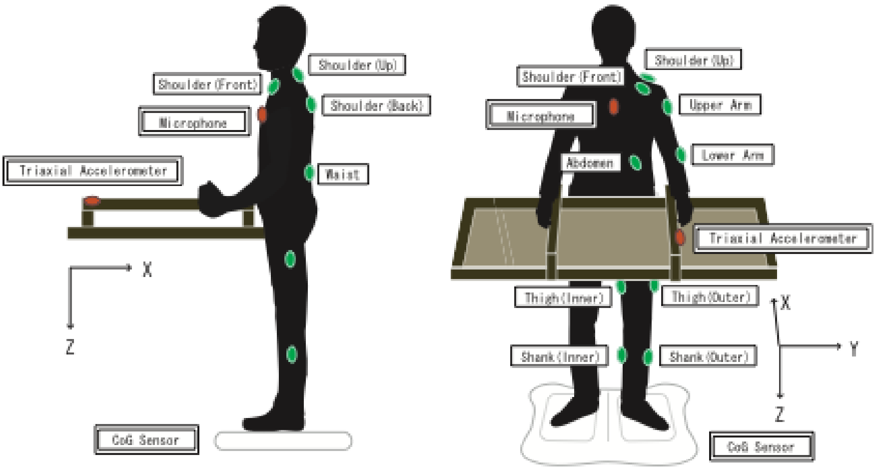


Fig. 2. Allocation of sensors

First of all, movies, CoG, sounds, EMG, acceleration are aligned on the viewer for analysis (Figure 3). As a result, acceleration of the tool showed significant periodicity (Figure 4). We further analyzed the data based on the periodicity of the movement.

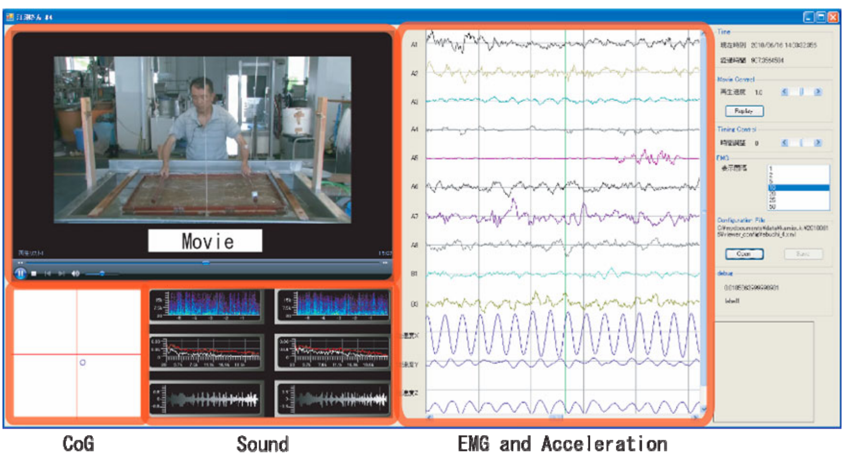


Fig. 3. Sensor data viewer for analysis

The periodicity of the movement of the tool was observed for both subjects. The periodicity is also associated with the movement of the body, which is shown in the EMG.

The result of subject A showed especially high periodicity. In addition, there were a sign of external force in the data of Z-axis acceleration. It is suggested that subject A uses the power of string of bamboo which support the tool (Figure 5).

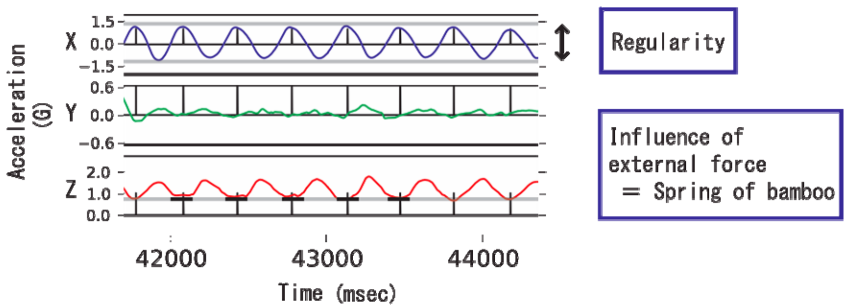


Fig. 4. Acceleration during X-axis swinging movement

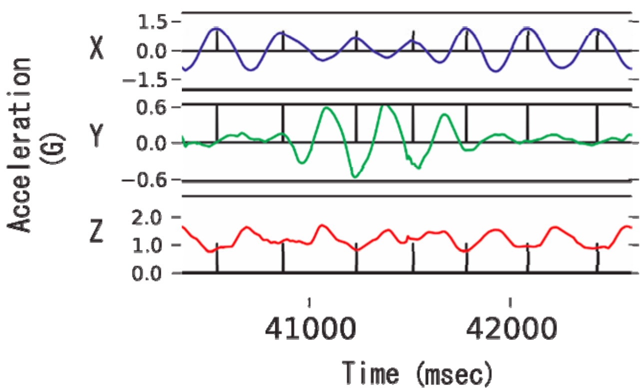


Fig. 5. Acceleration during X-Y combined swinging movement

5 Wearable Display System

From the result of measuring Kamisuki skills, remarkable periodicity have found in motion of tool and myoelectric signals. In traditional Japanese paper making, it is important to entwine paper fiber uniformly. Stable periodic motion will be one of the significant elements of its skills. According to previous researches result in induction of periodic motion by using motor-sense synchronization of human body in sound and vibration stimuli [9, 10]. We developed first-person wearable display system that induces periodic motion in Kamisuki by visual, audio and vibration stimuli. Figure 6 describes overall system. Vibration motors are placed on exact position where we measured myoelectric signals. Vibration stimuli are produced according to smoothed myoelectric potential data recorded in measuring experiment.

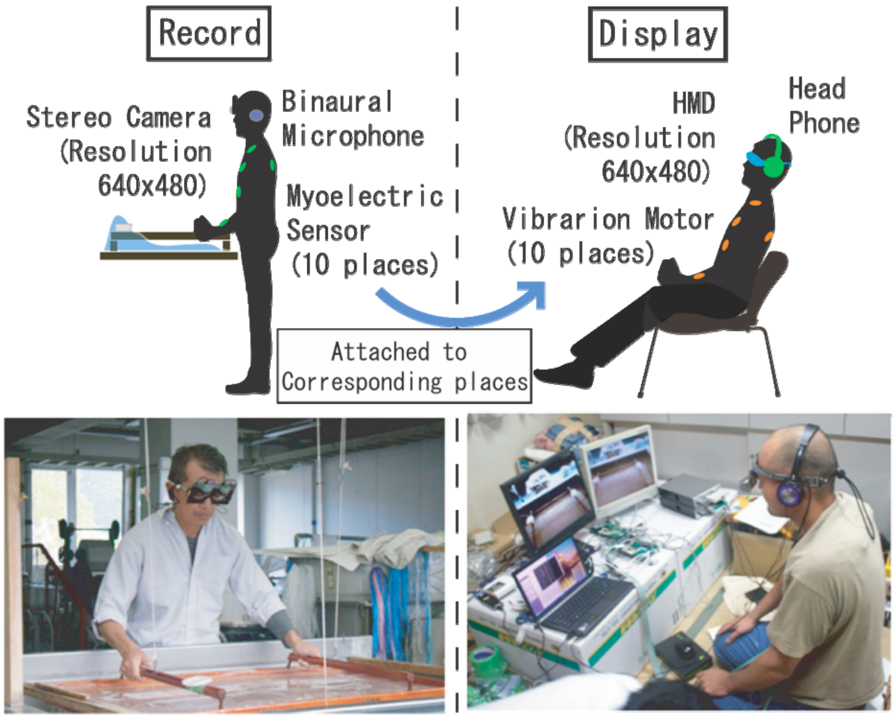


Fig. 6. Wearable recording system and displaying system

6 Experiment of Handing Down Kamisuki Skill

We conducted training experiment of developed wearable display system to 10 male subjects. One of the subjects is working as Kamisuki craftsman for ten years (subject B in skill measuring survey) and other subjects are all beginners. We asked subjects to observe subject A's (same subject in skill measuring survey) recorded first-person data through three different types of medium (wearable system with vibration stimuli, wearable system without vibration stimuli and video) and try to follow his motion in paper making process. Experiment procedure is shown in Figure 7.

Figure 8 indicates that each subject's paper making motion has similar and stable periodic motion to subject A if the data is plotted near the origin. Data in practicing period are not plotted since all the beginner subjects were not able to make stable periodic motion. Subject B could already perform similar motion cycle to subject A, therefore, there are no remarkable changes in subject B's data (see B1-B3 in Fig. 8). On the contrary, beginner subjects are more likely to produce similar periodic motion of subject A in the following order, wearable display with vibration stimuli (see F1-F2, D2 and C1 in Fig. 8), wearable display without vibration stimuli (see H1-H2, D1 and E3 in Fig. 6) and video (see J1-J2, C2 and E2 in Fig. 8). However, subjects who were not able to make periodic motion said that they could not figure out what to

Subjects	Learn I	Learn II	Learn III
B	△ Video only	□ Wearable B	○ Wearable A
C	○ Wearable A	△ Video only	–
D	□ Wearable B	○ Wearable A	–
E	○ Wearable A	△ Video only	□ Wearable B
F, G	○ Wearable A	○ Wearable A	–
H, I	□ Wearable B	□ Wearable B	–
J, K	△ Video only	△ Video only	–

Fig. 7. Training experiment procedure

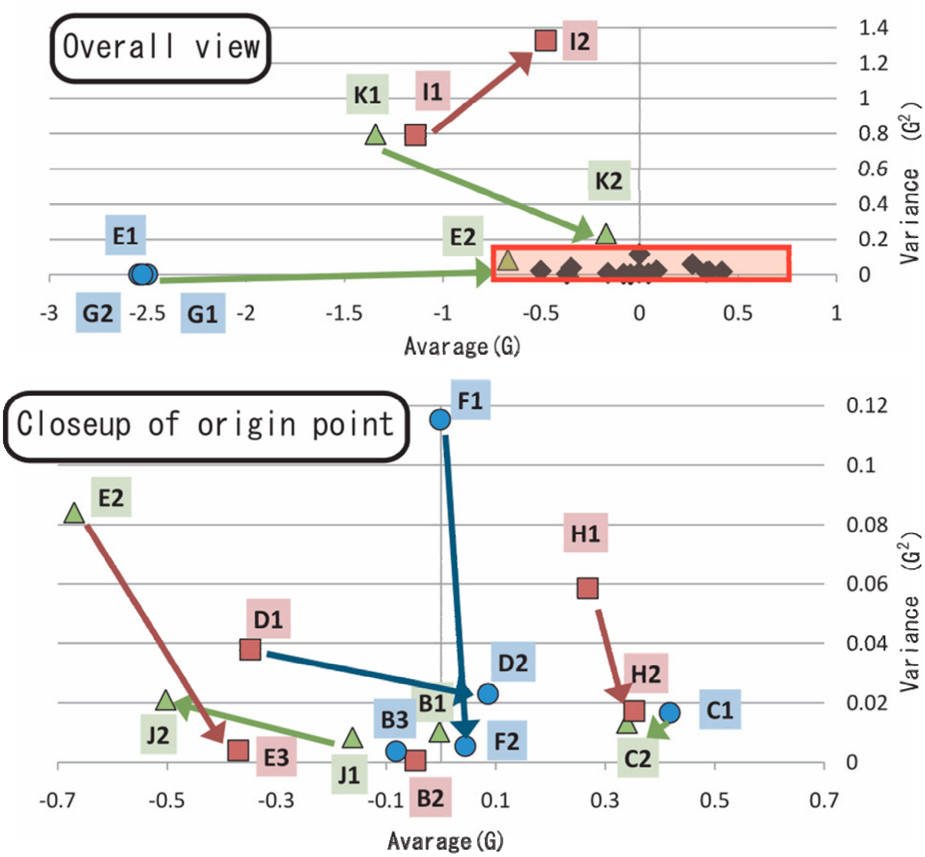


Fig. 8. Training experiment results

observe when experiencing wearable display systems (see K1-K2 and E1-E2 in Fig. 8). From the after experiment interview subjects who were able to make progress in producing periodic motion said that wearable display system can provide high immersive experience. Subject B said that he could observe the difference of his style of papermaking and subject A's style that is difficult to describe in words.

7 Conclusion

In this research, we measured craftsman's motion of Japanese traditional papermaking and found the stable periodic motion in trained craftsman. In order to handing down this basic periodic motion, we developed wearable display systems that provides first-person experience and induces learner of craftsman's motion in Japanese traditional papermaking. From the result of evaluation experiment, proposing wearable display system had the best performance in inducing craftsman's periodic motion. Wearable display with vibration stimuli has stronger effect in inducing periodic motion. Video observation can be helpful for leaner who could not catch up with multi modal information displayed through wearable display. Combination of using wearable display and video according to learner's level is suggested to be effective in learning of basic motion of papermaking.

Acknowledgments. This research is supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (A), 20240021.

References

1. Tanikawa, T., Ando, M., Wang, Y., Yoshida, K., Yamashita, J., Kuzuoka, H., Hirose, M.: A Case Study of Museum Exhibition - Historical Learning in Copan Ruins of Mayan Civiliza-tion. In: Proc. of IEEE Virtual Reality 2004, pp. 257–258 (2004)
2. Banno, A., Masuda, T., Oishi, T., Ikeuchi, K.: Flying Laser Range Sensor for Large-Scale Site-Modeling and Its Applications in Bayon Digital Archival Project. International Journal of Computer Vision 78(2-3), 207–222 (2007)
3. Koller, D., Turitzin, M., Levoy, M., Tarini, M., Croccia, G., Cignoni, P., Scopigno, R.: Protected interactive 3D graphics via remote rendering. In: Proc. of ACM SIGGRAPH, pp. 695–703 (2004)
4. Nakaoka, A., Nakazawa, K., Yokoi, H.: Generating whole body motions for a biped humanoid robot from captured human dances. In: Proc. of the IEEE ICRA, vol. 3, pp. 3905–3910 (2003)
5. Saga, S., Vlack, K., Kajimoto, H., Tachi, S.: Haptic video. In: ACM SIGGRAPH 2005 Emerging Technologies (2005)
6. Ito, R., Hiyama, A., Namiki, H., Miyashita, M., Tanikawa, T., Miyasako, M., Hirose, M.: Extraction and transmission of the biological implicit knowledge of artisan for skill acquisition. In: Proc. of ASIAGRAPH 2009, vol. 3(1), pp. 104–107 (2009)

7. Hiyama, A., Doyama, Y., Kakurai, K., Namiki, H., Miyasako, M., Hirose, M.: Archiving and Transferring of Traditional Artisanship focused on Interaction between an Artisan and Tools. In: Proc. of VSMM 2010 (2010)
8. All Japan Handmade Washi Association. Washi Handbook (2002) (in Japanese)
9. Muto, T., Miyake, Y.: Analysis of Co-emergence Process on the Human-Robot Cooperation for Walk-support. Transactions of the Society of Instrument and Control Engineers 38(3), 316–323 (2002) (in Japanese)
10. Watanabe, J., Ando, H., Maeda, T.: Shoe-shaped Interface for Inducing a Walking Cycle. In: Proceedings of the 15th International Conference on Artificial Reality and Telexistence, pp. 30–34 (2005)

Stroke-Based Semi-automatic Region of Interest Detection Algorithm for In-Situ Painting Recognition

Youngkyoon Jang and Woontack Woo

GIST U-VR Lab,
Gwangju 500-712, S. Korea
{yjang,wwoo}@gist.ac.kr

Abstract. In the case of illumination and view direction changes, the ability to accurately detect the Regions of Interest (ROI) is important for robust recognition. In this paper, we propose a stroke-based semi-automatic ROI detection algorithm using adaptive thresholding and a Hough-transform method for in-situ painting recognition. The proposed algorithm handles both simple and complicated texture painting cases by adaptively finding the threshold. It provides dominant edges by using the determined threshold, thereby enabling the Hough-transform method to succeed. Next, the proposed algorithm is easy to learn, as it only requires minimal participation from the user to draw a diagonal line from one end of the ROI to the other. Even though it requires a stroke to specify two vertex searching regions, it detects unspecified vertices by estimating probable vertex positions calculated by selecting appropriate lines comprising the predetected vertices. In this way, it accurately (1.16 error pixels) detects the painting region, even though a user sees the painting from the flank and gives inaccurate (4.53 error pixels) input points. Finally, the proposed algorithm provides for a fast processing time on mobile devices by adopting the Local Binary Pattern (LBP) method and normalizing the size of the detected ROI; the ROI image becomes smaller in terms of general code format for recognition, while preserving a high recognition accuracy (99.51%). As such, it is expected that this work can be used for a mobile gallery viewing system.[1]

Keywords: Semi-automatic ROI Detection, Hough-transform, Planar Object Recognition, Local Binary Pattern.

1 Introduction

Recent applications of mobile augmented reality (AR), such as ‘Layar’ [1] and ‘Wikitude’ [2], were based on points of interest (POI) focusing on ‘where’, when considering 5W1H. However, it does not provide specific information for an object instead of providing information of the area. Thus, recent research interests are now focusing on object-based recognition for mobile AR focusing ‘what’ inside of the ‘where’ as a specific type of context [3-5]. One recent application, ‘Google Goggles’, recognizes different kinds of objects including artwork and books [6]. However, it

[1] This research is supported by MCST and KOCCA, under the CT R&D Program 2011.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 167–176, 2011.

does not work properly in the case of altered illumination and severely skewed view directions for a less textured painting. A gallery, where many artworks are displayed, usually has similar challenging situations.

In order to recognize paintings inside of a gallery, Andreatta proposed appearance-based painting recognition for a mobile museum guide [7]. It uses a predefined center region of the captured image for the recognition. The method normalizes the image with the description grid of the region in order to overcome illumination changes of the environment. After the normalization, it detects overexposed areas (specular reflection) and uses the region with no noise. However, it does not support robust recognition when there are view direction changes, because it does not provide an ROI detection method. To overcome the challenge of skewed views and direction changes, the latest mobile guide was based on fast SIFT (Scale-Invariant Feature Transform) proposed by Ruf [8]. This approach utilizes currently available object recognition and network technologies. It captures an image from a mobile phone and sends it to the server. Then, the main procedure - feature extraction and matching - is done on server side. The feature-based approach generally provides accurate detection and recognition. However, it needs a great deal of processing time if it needs to recognize a painting's data due to severely a skewed view direction changes because it requires more training and comparison processes. Furthermore, the feature-based recognition does not provide accurate recognition results in terms of a less textured painting.

In order to overcome all the discussed problems and environment challenges of painting recognition, we propose stroke-based semi-automatic ROI detection algorithm for in-situ painting recognition. The proposed algorithm needs minimal participation by the user for specifying the initial search area of the painting's ROI by dragging a line from one corner of the painting to the other corner. Then, the proposed algorithm detects dominant edges by adaptively finding and using the threshold. Because the proposed algorithm adaptively finds the threshold, it handles both simple and complicated textured paintings and also detects dominant candidate lines of the ROI with Hough-transform method. Next, it estimates and detects unspecified vertices by selecting appropriate lines comprising the predetected vertices, even though it requires a stroke to specify two vertex searching regions (from one end of the ROI and the other diagonally) to begin. By taking the proposed semi-automatic (but simple) algorithm, it accurately detects the painting region, even if there is a severe view direction change and a user gives inaccurate input points. Moreover, the proposed algorithm normalizes the size of the detected ROI as smaller as the pre-determined size in order to make a general code format for recognition. Thus, the proposed algorithm provides robust recognition which is not affected by the various sizes of the paintings. Adopting the LBP [9] method has many advantages such as robust recognition to global illumination changes, fast processing time with binary code matching, and minimal storage size for saving codes. Experimental results of the proposed algorithm show high recognition accuracy (99.51%) with an accurate detection results (1.16 error pixels) for both view direction and illumination changes.

The rest of this paper is organized as follows. Section 2 introduces the general procedure of the proposed semi-automatic ROI detection algorithm, and a detailed description of the algorithm is presented in Section 3. Section 4 describes the implementation and experimental results of the proposed algorithm, and conclusions and future works are discussed in Section 5.

2 Overview of the Proposed Algorithm

An overview of the proposed algorithm can be seen in Figure 1. In order to initiate the algorithm, a user is required to capture a painting image from a mobile phone and drag a line specifying two initial search regions (red rectangular boxes in Figure 2(a)) on the touch screen of the phone like the yellow line of Figure 2(a). Then, the proposed algorithm determines the adaptive threshold corresponding to the captured painting and detects edges with the determined threshold in the specified regions. Next, it detects two vertices in both specified regions using the Hough transform. In this step, the algorithm gives line parameters in order to comprise the detected vertices. By taking one of two lines comprising one vertex, it finds the other line, which crosses inside an image. Then, it detects the ROI by estimating unspecified vertices (blue circle points in Figure 2(a)) and detecting accurate vertices (black blank circle points in Figure 2(a)) again inside of the estimated vertex regions.

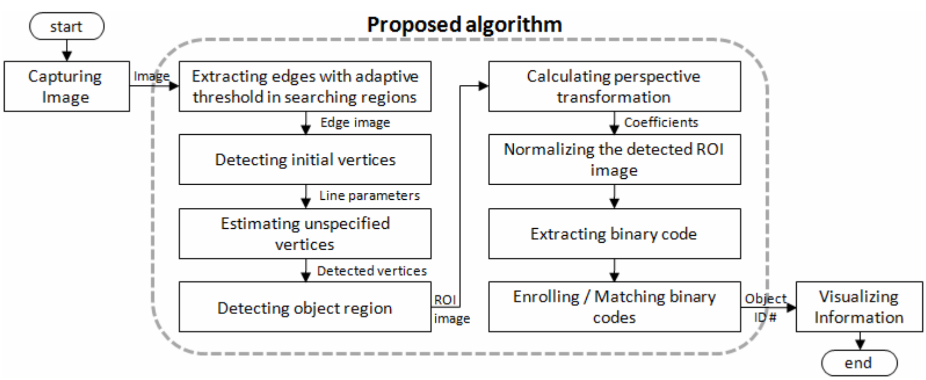


Fig. 1. Proposed algorithm flow

The detected ROI image is normalized by calculating perspective transformations, which affects the resultant general code size irrespective of the real size of the painting as shown in Figure 2(b). Figure 2(b) shows examples of the procedure for the binary code extraction and matching using the LBP method. By using the normalized ROI image, the proposed algorithm then extracts binary code via the LBP method [9], as shown in Figure 6. Next, the proposed algorithm enrolls or matches the extracted LBP code and preloaded LBP codes by calculating the Hamming Distance (HD) value. When the measured HD value is lower than the experimentally determined threshold (0.42), indicating a high level of similarity, the algorithm determines that the detected ROI image is a real painting, and vice versa. This proposed algorithm is suitable for mobile application and it provides fast and robust recognition in a challenging environment irrespective of view direction and illumination changes, even though the algorithm requires a user's minimal participation.

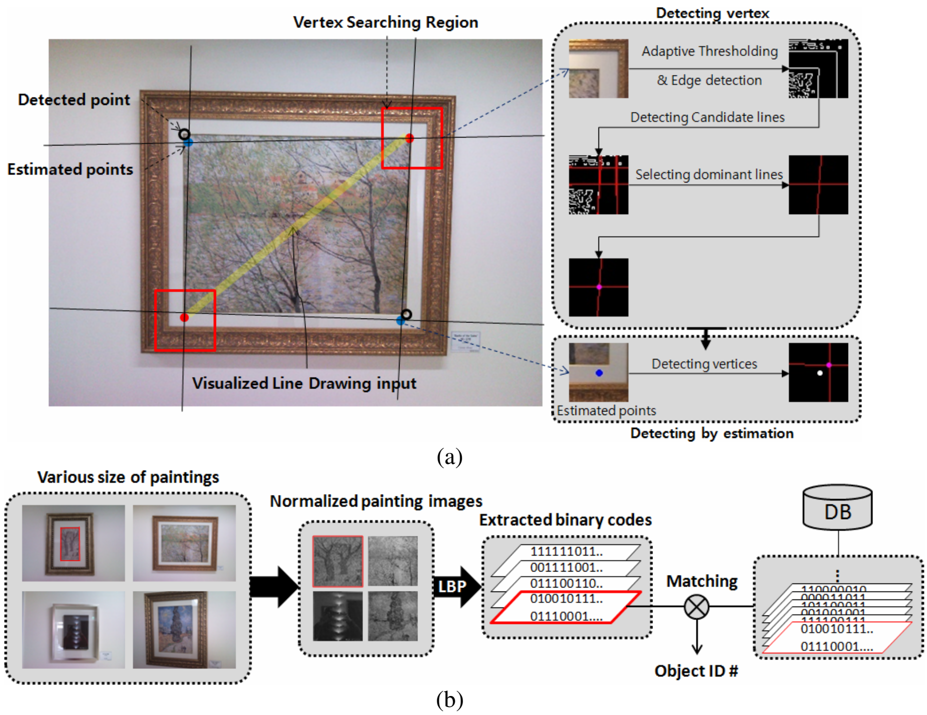


Fig. 2. Examples of the procedure for the proposed algorithm: (a) a stroke-based ROI detection (b) code extraction and matching

3 Proposed ROI Detection Algorithm for Robust Recognition

Because of the different texture amounts among paintings, the proposed algorithm needs to be able to extract dominant edges irrespective of the strength of the edges. Moreover, because the algorithm requires a user's participation (while preserving high detection and recognition accuracies despite view direction changes), it needs to be simple and to reduce the interaction process. To this end, we propose a stroke-based semi-automatic ROI detection algorithm using adaptively extracted edges and a vertex estimation process. Detailed explanations are provided as follows.

3.1 A Stroke-Based Semi-automatic ROI Detection with Vertex Estimation

To reduce the noise of the image in paintings with a large amount of texture, the general approach applies the Gaussian kernel to the captured image and extracts dominant edges. However, the approach does not extract weak edges such as blurred images or indistinct borders of the image, because it removes the weak edges from the borders by only using a predetermined threshold.

In order to overcome the problem, we first apply a global adaptive thresholding method only for the Intensity element of the captured image after converting the pixel format from RGB to HSV. Through the thresholding, we can extract dominant edges for the borders by using Canny edge detector as shown in Figure 3, because the adaptive thresholding method separates two regions. Even though the two regions have similar intensities (the paintings edges are not clearly defined) it automatically finds an appropriate threshold value. Therefore, it properly works for both simple and complicated textured paintings.

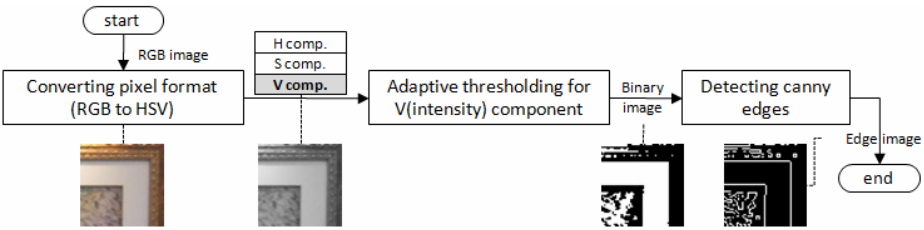


Fig. 3. Procedure of dominant edge detection

By using the detected dominant edges as an input, the Hough-transform detects dominant candidate lines by searching a limited range of parameters. Then, we select the two lines comprising one vertex. We do by examining the angle between two lines and the distance between the initially specified point and the cross point of the two lines. If both conditions are satisfied, as Equation (1), we detect the vertex as a result, as shown in Figure 4.

$$F(\rho_b,\theta_b,\rho_c,\theta_c)=\begin{cases} True, & MIN(D_d = Dist(P_s, P_d = Pt(\rho_b,\theta_b,\rho_c,\theta_c)), D_t)\ \&\ T_a < A = Ang(\rho_b,\theta_b,\rho_c,\theta_c) \\ False, & otherwise \end{cases}, \quad (1)$$

where (ρ_b, θ_b) are the parameters representing a base line, and (ρ_c, θ_c) are for a compared candidate line. We calculate intersection position P_d by using a $Pt()$ function with the four parameters. Then we check whether the distance D_d, between the detected point P_d and the specified point P_s is minimal compared with D_t. Shorter distance value is saved as D_t. Then, we also check that the angle, calculated by the function $Ang()$ with the parameters representing two lines, is larger than the pre-determined threshold T_a. If the two conditions are satisfied, function $F()$ returns true, meaning it was detected, and vice versa.

Like the procedure mentioned above, the algorithm detects two initial vertices inside of the user-specified searching regions. On the other hand, the other two vertices of the painting are estimated by using the two rho and theta values of the lines comprising the two detected vertices. For detecting the initial two vertices, we know the parameter values, rho and theta, of the detected lines through Hough transform. By taking one of those two lines comprising the detected one of those vertices, we can find the other line which crosses inside of an image, as shown in Figure 5 ①. Therefore, we detect object ROI by estimating the two other intersection points (pink dot in Figure 5 ②) as another two inputs for detecting vertices (white dot in Figure 5 ②). By comprising all the detected four vertices, we accurately detect ROI of a painting.

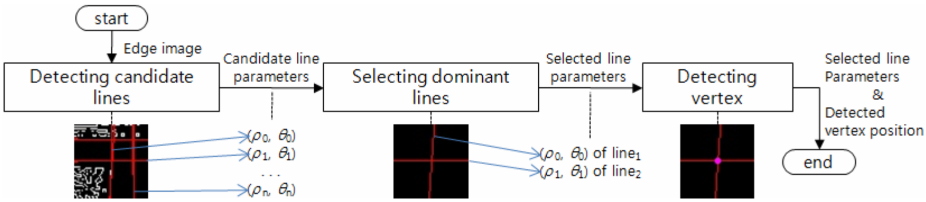


Fig. 4. Procedure of initial vertex detection

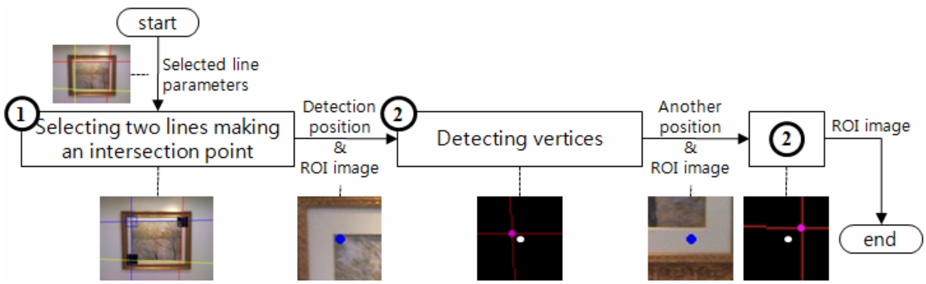


Fig. 5. Procedure of estimation (pink circle points) for detecting vertices at the unspecified position and its detection results (white circle points)

3.2 In-Situ Painting Recognition Based on Local Binary Pattern

To recognize both simple and complicated textured paintings in a challenging environment where the global illumination changes, it is important to extract illumination-invariant codes for the painting. For user's satisfaction the algorithm requires fast processing time even on mobile phones; however, the previous feature-based recognition [8] takes too much processing time on mobile phones. Furthermore, with less textured paintings, the previous method decreases the discrimination ability among the several less textured paintings because it does not extract enough numbers of the feature-points for the painting.

In order to overcome the problem, we simplified the LBP method [9] which considers the pattern of image intensity in the localized area. The original LBP method [9] considers rotation-invariant features. However, because the proposed algorithm has already detected the accurate ROI of the painting, the proposed algorithm only extracts the binary codes inside of down-sampled image (15×15) of the detected ROI. This is done by only adopting the idea of LBP code extraction. The

LBP can be defined as an ordered set of binary values determined by comparing the gray values of a center pixel and the eight neighborhood pixels around the center, as shown in Fig. 6. By using this simplified LBP method, we extract 225 bytes (8 * 15 * 15) of LBP codes representing a painting. Even though a user takes a painting image from the flank, the proposed algorithm corrects perspective distortion by applying the perspective transform. Then, it provides same size of the painting code and a simple matching algorithm with an Exclusive OR (XOR) binary operator as shown in Equation (2).

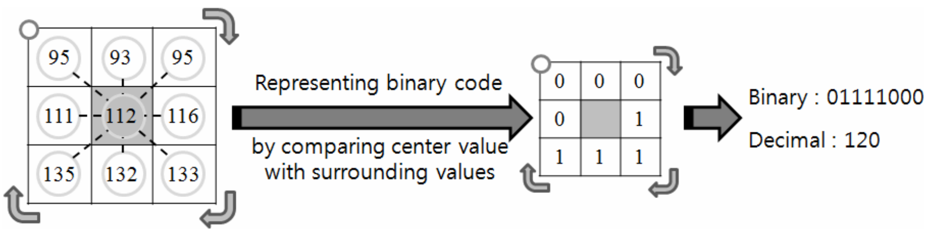


Fig. 6. LBP codes extraction

The currently extracted codes can be practically matched with the filtered set (number of the filtered image: 50) of codes among the code DB, because we assume that we know the direction and location of the user using sensors within the mobile phone. For matching, we use the Hamming distance measuring the dissimilarity between the current extracted code and the filtered set of codes, as represented in Equation (2). A smaller HD (Hamming distance) closer to 0 means the two codes are similar, and vice versa. We determined that the two codes are identical when the HD is less than the pre-determined threshold T_h. According to experimental results, we determined that T_h is 0.42, which shows the high recognition accuracy result.

$$HD = \frac{Num(codeA \otimes codeB)}{CodeBits}, \tag{2}$$

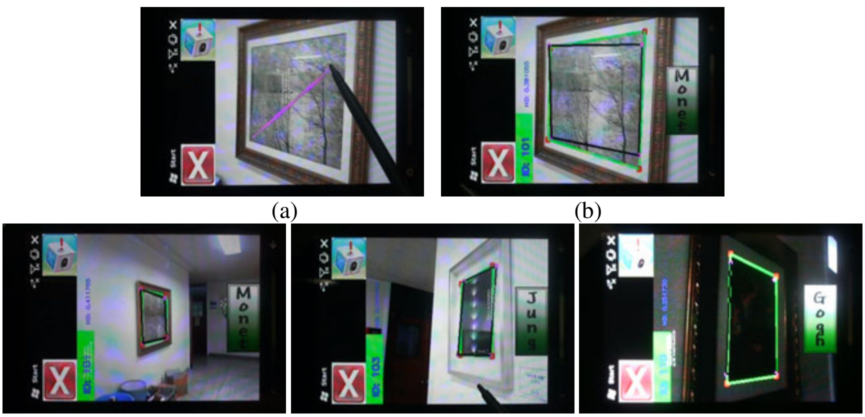
where CodeBits means the number of LBP code bits. In this paper, we extract 1800 (8 * 15 * 15) bits for LBP codes. $\otimes$ represents XOR operator. *Num*() function returns the number of 1's of the codes after the XOR operation. As it only uses the binary pattern for code matching, the proposed system using LBP-based recognition shows a fast processing time. Furthermore, the system could recognize less textured objects even in the case of illumination changes, because the pattern is extracted inside of the accurately detected ROI and it is globally illumination-invariant. Therefore, the proposed algorithm recognizes both less textured and more textured paintings in a challenging environment such as gallery even though the image is captured from the flank.

4 Implementation and Experimental Results

4.1 Implementation

In this work, we implemented a mobile painting recognition system prototype using the proposed a stroke-based ROI detection and LBP-based painting recognition algorithm. We used a DirectShow for capturing images from a camera of a mobile phone and an OpenCV library [10] for implementing basic image processing algorithms. We implemented our proposed algorithm on a mobile phone based on the Windows Mobile 6.3 platform. The phone had a 1 GHz ARMv7 processor and built-in (4.1 inch) touch-screen display.

We implemented the ROI detection algorithm as shown in Figure 7(b). The stroke interaction process is needed to specify the search areas as shown in Figure 7(a). The black line of Figure 7(b) shows an ROI comprised by the initially specified and estimated points. And the green line shows the detected ROI by using the proposed algorithm.



(c)

Fig. 7. Examples of the proposed algorithm's output: (a) user's dragging as an initial input; (b) ROI detection (black line: ROI consist of the initially specified and estimated points, green line: ROI comprised by the detected vertices); (c) results in the case of challenging environment (view distance, direction and illumination changes)

After the ROI is detected, it is normalized as shown in Figure 2(b). Then, the proposed algorithm extracts and matches the binary code. The recognized ID number corresponding to the painting and the artist's last name is on the left and right side of the screen.

As shown in Figure 7(b), the proposed algorithm detects an accurate ROI even though there was an inaccurate input. Moreover, the proposed algorithm is robust to the challenging environment (view distance, direction and illumination changes). Figure 7(c) presents the accurate detection and recognition results when the painting is captured from far distance (left), from the flank (middle) and in a dark environment (right). As such, it is expected that this work can be used for a mobile gallery viewing system.

4.2 Experimental Results

In this paper, we evaluated the performance of our proposed algorithm by using a database of captured gallery painting images. The database is comprised of 250 images acquired using a mobile phone images from 5 different angle under uncontrolled illumination conditions for 50 paintings. Each image in the database has a 24-bit color value and a pixel size of 320×240. This DB set is for experiments. In practice we only store LBP binary codes as a DB set for recognition.

First, we measured the processing time for the proposed algorithm. As a result, our proposed algorithm required 38.97 ms to run in our experimental environment. Specifically, ROI detection required 37.81 ms, the majority of the processing time, while the recognition required 1.16 ms including code extraction and matching, as shown in Table 1. Even without optimization of the source code, this speed is fast enough to get results in real-time on a mobile phone. With optimization, we could therefore expect even faster results during ROI detection.

Table 1. Processing time of the proposed algorithm

Algorithm	Processing time (ms)		
	Minimum	Maximum	Average
ROI detection	27.15	45.83	**37.81**
LBP code recognition	0.83	1.37	**1.16**
Entire processing time	27.98	47.20	**38.97**

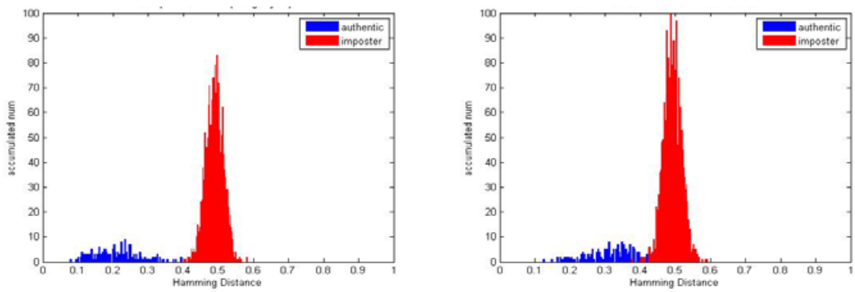
In the second set of tests, we measured the detection accuracies of the corners of the painting ROI, shown in Table 2. The accuracies were measured based on the Root Mean Square (RMS) pixel error between the detected points, and those chosen manually. As can be seen in Table 2, the average RMS pixel error indicates that the proposed algorithm detects accurate (1.16 error pixels) vertices comprising the ROI, even when users give an inaccurate input (4.53 error pixels, on average).

Table 2. Detection accuracies of the corners of painting ROI

Detection method	RMS pixel errors			
	Minimum	Maximum	Average	Standard. deviation
Manual selection	1.00	7.07	4.53	1.68
Proposed algorithm	**0.00**	**3.53**	**1.16**	**0.78**

Finally, we measured the recognition accuracy. For that we measured the Equal Error Rate (EER), which represents the minimum error rate for recognition by taking an exact threshold of the given test set. As experimental results, the proposed detection algorithm supports accurate recognition (EER: 0.48%), whereas the EER was 3.02% without the algorithm (manual selection), as shown in Figure 8.

Thus, we could confirm that the proposed algorithm guarantees robust painting recognition, even though a user may specify inaccurate inputs and in cases in which there are view direction angle and illumination changes.



(a) (b)

Fig. 8. Experimental results for recognition: (a) with the proposed detection algorithm, and (b) without the proposed detection algorithm

5 Conclusions and Future Works

This paper proposed a stroke-based ROI detection algorithm for in-situ painting recognition. We plan to develop an automatic detection algorithm for convenient recognition, while maintaining high recognition rates.

References

1. Layar, `http://www.layar.com/`
2. Wikitude, `http://www.wikitude.org/`
3. Lee, Y., Oh, S., Shin, C., Woo, W.: Ubiquitous Virtual Reality and Its Key Dimension. In: International Workshop on Ubiqiutous Virtual Reality 2009, pp. 5–8 (2009)
4. Shin, C., Kim, H., Kang, C., Jang, Y., Choi, A., Woo, W.: Unified Context-aware Augmented Application Framework for Supporting User-Driven Mobile Tour Guide. In: 8th International Symposium on Ubiquitous Virtual Reality 2010, pp. 52–55 (2010)
5. Kim, H., Woo, W.: Real and Virtual Worlds Linkage through Cloud-Mobile Convergence. In: Workshop on Cloud-Mobile Convergence for Virtual Reality, pp. 10–13 (2010)
6. Google Goggles, `http://www.google.com/mobile/goggles/`
7. Andreatta, C., Leonardi, F.: Appearance Based Paintings Recognition For a Mobile Museum Guide. In: International Conference on Computer Vision Theory and Applications, VISAPP 2006 (2006)
8. Ruf, B., Kokiopoulou, E., Detyniecki, M.: Mobile Museum Guide Based on Fast SIFT Recognition. In: 6th International Workshop on Adaptive Multimedia Retrieval (2008)
9. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution Gray Scale and Rotation Invariant Texture Analysis with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 24(7), 971–987 (2002)
10. Intel Open Source Computer Vision Library, `http://sourceforge.net/projects/opencvlibrary/`

Personalized Voice Assignment Techniques for Synchronized Scenario Speech Output in Entertainment Systems

Shin-ichi Kawamoto[1,3], Tatsuo Yotsukura[2], Satoshi Nakamura[3], and Shigeo Morishima[4]

[1] Japan Advanced Institute of Science and Technology,
1-1 Asahidai, Nomi, Ishikawa, 923-1211, Japan
kawamoto@jaist.ac.jp
[2] OLM Digital Inc.,
Mikami Bldg. 2F, 1-18-10 Wakabayashi, Setagaya-ku, Tokyo, 154-0023, Japan
yotsu@olm.co.jp
[3] National Institute of Information and Communications Technology,
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0289, Japan
satoshi.nakamura@nict.go.jp
[4] Waseda University,
3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan
shigeo@waseda.jp

Abstract. The paper describes voice assignment techniques for synchronized scenario speech output in an instant casting movie system that enables anyone to be a movie star using his or her own voice and face. Two prototype systems were implemented, and both systems worked well for various participants, ranging from children to the elderly.

Keywords: Instant casting movie system, post-recording, speaker similarity, voice morphing, synchronized speech output.

1 Introduction

Instant casting movie system (ICS) is an entertainment system that allows participants to instantly appear in a movie as a CG character [1]. In ICS, all processes are performed automatically (Fig. 1): scanning of the face shape and image, reconstruction of the 3D face model, and generation of onscreen appearance and movement. However, assigned voices of the participants' CG characters were only switched pre-recorded male or female voices, depending on the gender of the participant. Our goal is to extend the functionality of voice assignment for the ICS. The participant's CG character should be assigned a voice that is matched and similar to the participant's own voice. In addition, the voice quality should be better matched to the CG movie quality. Since the CG character has various facial expressions, the voice should communicate various expressions of emotions. Various speech synthesis techniques have been proposed [2-5]. However, it is difficult to generate high quality voices that match the various scenes.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 177–186, 2011.

Excessive deformation of the speech waveform is especially difficult without degrading voice quality.

In this paper, we introduce three approaches to extend the voice assignment function: quick post-recording, similar speaker selection and voice morphing. For short scenarios, participants can record scenario speeches directly using our post-recording tool. For long scenarios, participants can participate in the movie easily using the combined approaches of similar speaker selection and voice morphing. We have developed prototype speech synchronization systems. Scenario speeches can be output synchronously along with the CG character using our systems.

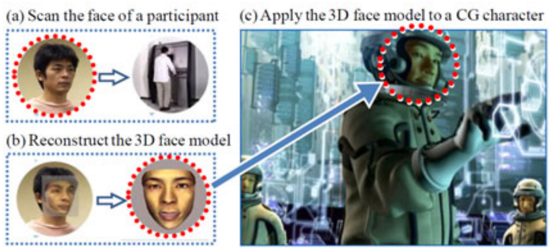


Fig. 1. Instant casting movie system (ICS)

2 Quick Post-recording Tool

If the most significant feature of voice casting is that the CG character acts as the participant with the participant's voice, one of the best solutions is to directly record the scenario voice from the participants. However, it is difficult for participants to record their own voices for synchronization with the movie. This recording task, called post-recording, is widely used in movie and animation production. The accuracy of synchronization between voice and acting depends on the skill of voice actors. Adjusting the timing of the scenario speech takes a long time for non-professional voice actors. In addition, these post-recording environments for professional use are expensive and complicated, and must be manually operated by skillful sound engineers. Considering the cooperative work of ICS, the following system requirements of the post-recording tool must be provided: 1) various timings of post-recording information; 2) automatic post-processing function of recorded voices.

2.1 Designing the Timing Information of Post-recordings

An important system requirement of a post-recording system is to produce scenario speech synchronously with the character's action in the movie. To fulfill this requirement, our tool provides users six types of timing information: movie, BGM/SE, time code, reference voice for voice-over acting, colored text like karaoke, and rhythmic information. Rhythmic information is generated by repetitively displaying the target scene over a constant period. Our tool can also be selected with/without a reference voice for voice-over acting, since the reference voice sometimes influences the user's acting.

2.2 Automatic Post-processing

To use our post-recording tool in ICS, participant's voices should be inserted into a movie soon after they are recorded. However the post-processing of recorded voices is time-consuming, especially for sound engineers who manually extract the essential parts of recorded voices to be synchronized with the corresponding movie shots. Thus an automatic post-processing mechanism for recorded voices should be implemented to eliminate such time-consuming work.

2.3 Prototype of Post-recording Tool

We developed a post-recording tool that fulfills the requirements. Fig. 2 shows a screenshot of our tool. Participants operate this tool with a touch-screen. The user speaks the script, which is synchronized with the video images. This tool has the following functions for supporting voice and lip synchronization: displaying the time code and the transcript text, and voice-over enable/disable buttons. The voice volume of user utterances can be checked using a meter. The video images for the post recording are repeatedly played by pushing the "OK" button to send a signal that the user felt a sentence recorded correctly. In our tool for ICS, manual post-processing is not needed, since our tool automatically extracts and saves the user voices that fit the length of the target scenes.

Fig. 2. Screenshot of post-recording tool

2.4 Evaluation of Post-recording Tool

In this section, we evaluate our prototype system in terms of three points: stability of real world use, ease of operation, and shortening of voice acquisition time.

Stability. We exhibited our system from February 10th to 11th, 2008 at the Miraikan Museum in Tokyo. The system worked well with 162 participants (including 79 children) for post-recording. While the children needed the support of a parent or an attendant to use our system, many participants could easily understand and operate it.

Ease of operation. We ran a subjective test to evaluate our system. 99 subjects between the ages of 18 and 63 tested it and answered some questions. To evaluate ease of operation, we asked the following question: "*Was the operation of this post-recording system easy?*" As a result. 64% of the subjects responded that our system's interface could be operated easily (Very Easy=13%, Easy=51%, Neither=11%, Difficult=11%, Very Difficult=23%).

Acquisition time. To evaluate the efficiency of data acquisition, we measured operation time using our system. For comparison, we measured the manual recording operation time using popular sound recording equipment with an operator. In manual recording, the subject speaks the lines of the script while watching the movie and the scenario text, and listening to the BGM/SE and the reference voice. After this manual recording, the operator must manually split the recorded voice using sound editing software. In manual recording, the average recording and splitting operation times of the eleven subjects were 64 minutes and 69 minutes, respectively. On the other hand, using our system, the recording operation time averaged 77 minutes with 91 subjects. In our system, the voice splitting operation time includes the recording time. Our system reduced the total recording time by 42% in comparison with manual recording. In addition, the number of professional operators can also be reduced using our system, since the subjects themselves operated our system in this evaluation of operating time.

3 Selecting Similar Speakers

The similar speaker selection process selects from the voice actor DB an actor with a voice similar to that of the participant, and assigns the selected voice to the character. In this approach, the voice quality does not undergo any degradation. However, the possibilities with regard to the extent of voice similarity that can be achieved are limited to the size of the voice actor DB. Various acoustic features related to the perception of speaker similarity were reported independently [6-9]. However the personality of a speaker is evinced not only in the voice quality but also in the prosodic intonation. Therefore, we need to consider multiple acoustic features for various voice characteristics. The key technology of our method is to combine multiple acoustic features that are used to calculate perceptual similarity into our system implementation so as to realize within the ICS, a closely-matched voice for each participant's character.

3.1 Estimation Method

We estimate the perceived similarity of speakers by participants using a combination of multiple acoustic features for greater accuracy. The perceptual similarity estimate s is calculated using Equation 1.

$$s = -\sum_{i=1}^{n} \alpha_i x_i \qquad (1)$$

In this equation, n is the number of acoustic features, x_i is the distance of the ith acoustic feature between the voices, and α is the weighting coefficient for each such distance between acoustic features.

We use 8 acoustic features related to voice personality. These are the Mel Frequency Cepstral Coefficient (Static: 12 + Dynamic: 13 = 25 dimensions)[6], the STRAIGHT Cepstrum of over 35 dimensions and 1st dimension [7], the Spectrum of over 2.6 kHz, the STRAIGHT-Ap under 2 kHz [8] that is a parameter of STRAIGHT [10], the fundamental frequency, the formants (F1 - F4), and the spectrum slope between 0 kHz - 3 kHz [9]. To extract these acoustic features, we use a window length of 25 ms and the shift rate of 10ms.

We calculate the distance between the acoustic features with Dynamic Time Warping (DTW). DTW distance is commonly used in a wide range of pattern recognition systems. It can estimate perceptual similarities accurately because it represents the temporal structure of acoustic features.

3.2 Optimization of Weighting Coefficients

To increase the correlation between the perceived similarity as reported by the subjects and that estimated using our method, we optimize the weighting coefficient α in Equation 1. To select a target speaker from a speaker DB, we represent the perceived similarities of the other speakers to the target by ranking the speakers in a permutation. The ranking is determined by quick sort based on subjective judgment. A subject judges similarity considering various speech features. Then the weighting coefficients α are optimized using the steepest descent method to increase the Spearman's rank correlation coefficient between the ranking of perceived similarity and that of acoustic similarity. Acoustic similarity is calculated using Equation 1. Spearman's rank correlation is shown in Equation 2.

$$\rho = 1 - \frac{6\sum_{i=1}^{n} (\beta_i - \gamma_i)^2}{N^3 - N} \qquad (2)$$

In this equation, β is the ranking with respect to perceived similarity ascribed by the subject. γ is the acoustic similarity ranking derived by using our method. N is the number of units of speech data. For this optimization, we used speech data uttered by 36 speakers.

4 Voice Morphing

The Voice morphing approach is based on blending a few voices to generate a voice similar to that of the participant. Our approach is based on STRAIGHT [10] voice morphing, which is an extra-high-quality voice morphing technique [11]. The key technology in our approach enables the automatic estimation of the optimal blending weights required to generate a voice similar to that of the participant.

4.1 Two Speakers' Voice Morphing

The basic idea of voice morphing is to generate an intermediate voice from two source voices by using an arbitrary blending ratio [12]. STRAIGHT-based morphing [11] handles the feature vectors of time-frequency representation derived by STRAIGHT [10], STRAIGHT spectrogram, Aperiodicity Map and F0. Time-frequency transformation of each feature vector is represented as a simple piecewise bilinear transformation with the same blending ratio.

4.2 Multiple Speakers' Voice Morphing

Takahashi et al. extended a conventional STRAIGHT-based morphing system to a multiple-speaker morphing mechanism [13]. The procedure is almost the same as that of conventional STRAIGHT-based morphing; it involves 1) anchor points, characteristic corresponding points in the time-frequency domain that are manually assigned on each reference spectrogram; 2) time-frequency transformation, which is derived from target and reference feature vectors based on the anchor points; and 3) reference feature vectors, which are morphed to mapped target feature vectors $\boldsymbol{x}_{mrp}$ based on Equation 3 with a blending ratio vector $\boldsymbol{r}$.

$$x_{mrp} = \sum_{s=1}^{S} r_s \boldsymbol{x}_s \tag{3}$$

where r_s is the blending ratio for speaker s, and $\boldsymbol{x}_s$ are the feature vectors for speaker s.

4.3 Voice Morphing for Generating Specific Speakers

We introduce our technique which can estimate a blending ratio vector to generate a specific speaker's voice based on multiple-speaker morphing. In this paper, we estimate a blending ratio vector that satisfies the following formula.

$$\hat{\mathbf{r}} = \arg\min_{r} \left\| \mathbf{y} - \hat{\mathbf{x}}_{mrp} \right\|^2 = \arg\min_{r} \left\| \mathbf{y} - \sum_{s=1}^{S} r_s \mathbf{x}_s \right\|^2 \tag{4}$$

$$\mathbf{r} = [r_1, r_2, \ldots, r_S]^{\mathrm{T}} \tag{5}$$

where $\mathbf{y}$ is the feature vector of target speaker, and $\hat{\mathbf{r}}$ is the estimated blending ratio vector. In this method, the blending ratio vector is minimized to yield the following formula.

$$e(\mathbf{r}) = \sum_{\tilde{f}=1}^{F} \sum_{t=1}^{\tilde{T}(\mathbf{r})} \left(y_{\tilde{t}}(f) - x_{\tilde{t}}(\tilde{f})\mathbf{r} \right)^2 \tag{6}$$

where $y_{\tilde{t}}(f)$ is the feature vector of a target speaker with a regularized time domain, $x_{\tilde{t}}(\tilde{f}) = \left[x_{\tilde{t}}^{(1)}(\tilde{f}), x_{\tilde{t}}^{(2)}(\tilde{f}),..., x_{\tilde{t}}^{(S)}(\tilde{f})\right]$ are feature vectors for reference speakers, and S is the number of reference speakers. $\tilde{f}$ and $\tilde{t}$ refer to time and frequency domain elements, respectively, regularized by anchor points and a blending ratio vector $\hat{\mathbf{r}}$. $\tilde{T}(\hat{\mathbf{r}})$ is the regularized speech duration in the time domain, as shown below.

$$\tilde{T}(\hat{\mathbf{r}}) = \sum_{s=1}^{S} \hat{\mathbf{r}}_s T_s \tag{7}$$

Since the STRAIGHT-based voice morphing process controls various features by the same blending ratio vector in the time-frequency domain, it is difficult to solve Equation 6 analytically. Therefore, we use an iterative approach to solve for a blending ratio vector using the following formulae.

$$\hat{\mathbf{r}}_{n+1} = \hat{\mathbf{r}}_n - \alpha E_n \tag{8}$$

$$E_n = \left(\frac{\partial^2 e'(\hat{\mathbf{r}}_n)}{\partial \hat{\mathbf{r}}_n{}^2}\right)^{-1} \frac{\partial e'(\hat{\mathbf{r}}_n)}{\partial \hat{\mathbf{r}}_n} = \left(\overline{\mathbf{X}}_n{}^T \overline{\mathbf{X}}_n\right)^{-1} \overline{\mathbf{X}}_n{}^T \left(\overline{\mathbf{Y}}_n - \overline{\mathbf{X}}_n \hat{\mathbf{r}}_n\right) \tag{9}$$

where $e'(\hat{\mathbf{r}})$ is an approximation formula of Equation 6 that assumes the blending ratio $\hat{\mathbf{r}}_n$ is constant. $\overline{\mathbf{X}} = \left[\overline{\mathbf{X}}_1{}^T, \overline{\mathbf{X}}_2{}^T,..., \overline{\mathbf{X}}_{\tilde{T}(\hat{\mathbf{r}}_n)}^T\right]^T$ are regularized feature vectors in the time-frequency domain that are updated n times by the blending ratio vector $\hat{\mathbf{r}}_n$. $\overline{\mathbf{Y}} = \left[\mathbf{y}_1{}^T, \mathbf{y}_2{}^T,..., \mathbf{y}_{\tilde{T}(\hat{\mathbf{r}}_n)}^T\right]^T$ are regularized feature vectors in the time domain that are updated by a blending ratio vector $\hat{\mathbf{r}}_n$. $X_{\tilde{t}_n}$, $\mathbf{x}_{\tilde{t}_n}^{(s)}$, and $\mathbf{y}_{\tilde{t}_n}$ are defined as follows.

$$\mathbf{X}_{\tilde{t}_n} = \left[\mathbf{x}_{\tilde{t}_n}^{(1)}, \mathbf{x}_{\tilde{t}_n}^{(2)},..., \mathbf{x}_{\tilde{t}_n}^{(S)}\right] \tag{10}$$

$$\mathbf{x}_{\tilde{t}_n}^{(s)} = \left[x_{\tilde{t}_n}^{(s)}(1), x_{\tilde{t}_n}^{(s)}(2),..., x_{\tilde{t}_n}^{(s)}(\tilde{f}_n),..., x_{\tilde{t}_n}^{(s)}(F)\right]^T \tag{11}$$

$$\mathbf{y}_{\tilde{t}_n} = \left[y_{\tilde{t}_n}(1), y_{\tilde{t}_n}(2),..., y_{\tilde{t}_n}(\tilde{f}_n),..., y_{\tilde{t}_n}(F)\right]^T \tag{12}$$

In this paper, we adopt α as 1, the number of iterations as 20, and the number of reference speakers as 8.

5 Implementation of Synchronized Speech Output

Fig. 3 shows our prototype system for ICS. Participants record their own speech using recording PCs. Input is a voice speaking a sentence which was recorded by each participant. Similar Voice Selection Servers and Voice Morphing Servers calculate the results of scenario speeches that are intended to be similar to the participants' voices. We have recorded 60 different kinds of voices to construct the voice DB of this system. This DB covers a wide range of participants, in terms of age (subjective ages: 6-63). In addition, this DB has a balanced male-female ratio (subjective gender: male=29, female=31). These scenario speeches are played based on a Longitudinal Time Code (LTC) that represents time synchronization with video images. The Display PC outputs video images and stereo audio, which consists of the LTC and the recorded sound (a mixture of BGM and SE). The audio composite PC sends speech data to the mixer based on the LTC. The mixture of speech data and sound is sent to the audio speakers. The image composite PC also sends video data to the display based on the LTC. As for the prototype system's movie content, we used “Grand Odyssey,” which was exhibited at the 2005 World Exposition in Aichi, Japan. Our system can be used easily and quickly by various participants-ranging from children to the elderly-because it is based on participants having to record one specific sentence only. We exhibited our system on March 20-22, 2009, at the Miraikan Museum in Tokyo. The system was tested by over 100 participants, including children and elderly people, and was found to work well.

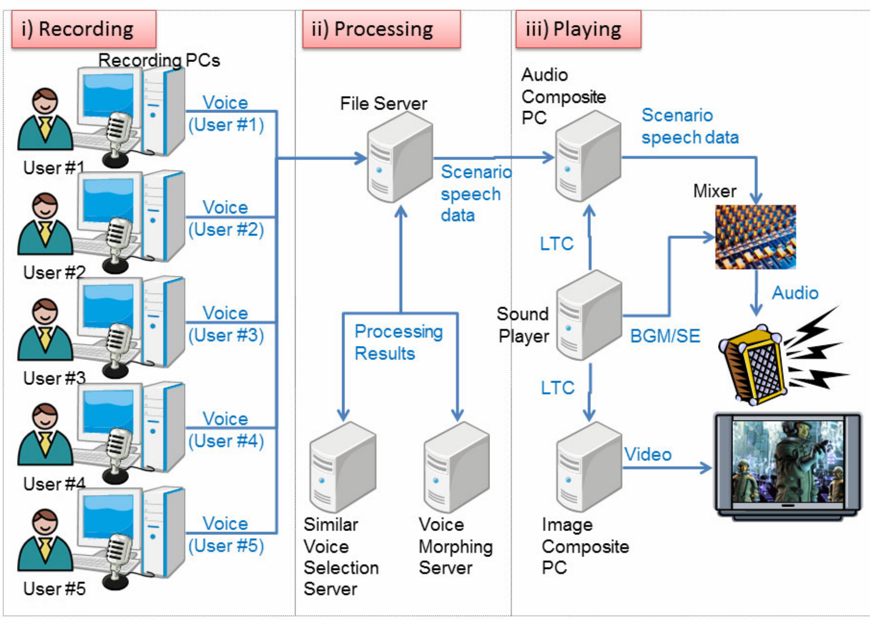


Fig. 3. Overview of prototype system

6 Discussion and Conclusion

In this paper, we described two types of scenario speech assignment systems to extend the casting functions of ICS. Our system worked well for various participants. Since synchronization of these scenario speeches was based on LTC, all three approaches can be combined into the same system.

The voice post-recording tool can be operated intuitively with various timing information. In this system, participants feel that they participate directly in the movie production, since participants uttered their scenario speech directly. However, the post-recording task is time-consuming, if the scenario speech is too long. In our post-recording strategy, scenario speech should be kept short to avoid overload of participants' tasks. In addition, it is difficult for most people to utter exaggerated expressions matched to the target scenes. Thus, the quality of movie output depends on participants. One of our future target projects is to modify the expression style of recorded voices automatically to match the target scenes.

The combined approach of similar voice selection and voice morphing also worked well by recording an input sentence for each participant. In this approach, workload of participants is less than in the post-recording approach, since participants only recorded a read speech, without exaggerated expression. Quality of output movie is stable, since scenario speech is based on voice DB, which was recorded by professional voice actors preliminarily. However, similarity of output voice depends on the size of voice DB. It is important to establish an archetype for the design strategy for the construction of a voice actor DB. Constructing a voice actor DB is a time- and money-intensive task. In voice morphing, assigning anchor points is also time-consuming. In order to improve the efficiency of these tasks, our system needs to incorporate other features. In addition, the use of the voice morphing technique leads to a slight degeneration in the speech quality of output voices. At present, we check the voice quality of outputs manually. One important course of future work is to develop an automatic speech quality evaluation technique. This technology will reduce operational costs of our system.

Acknowledgments. This research is supported by the Special Coordination Funds for Promoting Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology.

References

1. Maejima, A., Wemler, S., Machida, T., Takebayahashi, M., Morishima, S.: Instant Casting Movie Theater: The Future Cast System. The IEICE Transactions on Information and Systems E91-D(4), 1135–1148 (2008)
2. Zen, H., Nose, T., Yamagishi, J., Sako, S., Masuko, T., Black, A.W., Tokuda, K.: The HMM-based speech synthesis system version 2.0. In: Proc. of ISCA SSW6, Bonn, Germany (2007)
3. Kawai, H., Toda, T., Yamagishi, J., Hirai, T., Ni, J., Nishizawa, N., Tsuzaki, M., Tokuda, K.: XIMERA: A Concatenative Speech Synthesis System with Large Scale Corpora. IEICE Trans. J89-D-II(12), 2688–2698 (2006)

4. Hunt, A., Black, A.: Unit selection in a concatenative speech synthesis system using a large speech database. In: Proc. of ICASSP, pp. 373–376 (1996)
5. Clark, R.A.K., Richmond, K., King, S.: Multisyn: Open-domain unit selection for the Festival speech synthesis system. Speech Communication 49(4), 317–330 (2007)
6. Reynolds, D.: Robust text-independent speaker identication using gaussian mixture speaker models. IEEE Trans. On Acoust. Speech and Audio Processing 3(1) (1995)
7. Kitamura, T., Saitou, T.: Contribution of acoustic features of sustained vowels on perception of speaker characteristic. In: Proc. of Acoustical Society of Japan 2007 Spring Meeting, pp. 443–444 (2007)
8. Saitou, T., Kitamura, T.: Factors in /vvv/ concatenated vowels affecting perception of speaker individuality. In: Proc. of Acoustical Society of Japan 2007 Spring Meeting, pp. 441–442 (2007)
9. Higuchi, N., Hashimoto, M.: Analysis of acoustic features affecting speaker identification. In: Proc. of EUROSPEECH, pp. 435–438 (1995)
10. Kawahara, H.: Straight: An extremely high-quality vocoder for auditory and speech perception research. In: Greenberg, Slaney (eds.) Computational Models of Auditory Function, pp. 343–354 (2001)
11. Kawahara, H., Matsui, H.: Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. In: Proc. of ICASSP, vol. 1, pp. 256–259 (2003)
12. Slaney, M., Covell, M., Lassiter, B.: Automatic audio morphing. In: Proc. of ICASSP, pp. 1001–1004 (1995)
13. Takahashi, T., Nishi, M., Irino, T., Kawahara, H.: Average voice synthesis using multiple speech morphing. In: Proc. of Acoustical Society of Japan 2006 Spring Meeting, pp. 229–230 (2006)

Instant Movie Casting with Personality: Dive into the Movie System

Shigeo Morishima[1], Yasushi Yagi[2], and Satoshi Nakamura[3]

[1] Dept. of Advanced Science and Engineering, Waseda University, 3-4-1 Okubo Shinjuku-ku, Tokyo, Japan
[2] The Institute of Scientific and Industrial Research, Osaka University 8-1 Mihogaoka, Ibaraki, Osaka, Japan
[3] National Institute of Information and Communications Technology, 3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan
shigeo@waseda.jp, yagi@am.sanken.osaka-u.ac.jp, satoshi.nakamura@nict.go.jp

Abstract. "Dive into the Movie (DIM)" is a name of project to aim to realize a world innovative entertainment system which can provide an immersion experience into the story by giving a chance to audience to share an impression with his family or friends by watching a movie in which all audience can participate in the story as movie casts. To realize this system, we are trying to model and capture the personal characteristics instantly and precisely in face, body, gait, hair and voice. All of the modeling, character synthesis, rendering and compositing processes have to be performed on real-time without any manual operation. In this paper, a novel entertainment system, Future Cast System (FCS), is introduced as a prototype of DIM. The first experimental trial demonstration of FCS was performed at the World Exposition 2005 in which 1,630,000 people have experienced this event during 6 months. And finally up-to-date DIM system to realize more realistic sensation is introduced.

Keywords: Personality Modeling, Gait Motion, Entertainment, Face Capture.

1 Introduction

The purpose of DIM project is to aim to realize an innovative entertainment system in which especially all of the audience can become a movie cast instantly and can enjoy an immersive feeling and excited impression in the story. Future Cast System is one of a prototype of DIM which first was exhibited at the 2005 World Exposition in Aichi Japan. The Future Cast System offers two key features. First, it provides each member of the audience with the amazing experience of participating in a movie in real time. Second, the system can automatically perform all of the processes required to make this happen: from capturing the shape of each audience member's face, to generating a corresponding CG face in the movie. To create a CG character which closely resembles the visitor's face with FCS system, an original 3D range scanner first acquires the 3D geometry and texture of the visitor's face. Second, facial feature points, such as eyes, eyebrows, nose, mouth, and the outline of the face are extracted from the frontal face image. Next, a generic facial mesh is adjusted to smoothly fit

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 187–196, 2011.

over the feature points of the visitor's face. Scanned depth of the visitor's face which corresponds to vertices on the deformed generic facial mesh is acquired and then the visitor's frontal face image is attached to the mesh. The entire process, from the face scanning stage to the CG character generation stage, is performed automatically as soon as possible before the show starts. Then, scanned data, including the face model, estimated age and gender information are transferred to a file server.

At the time of the show, we call up the visitor's character information, such as facial geometry, texture, and so on, from the file server and control the face using Cast/Environment Scenario Data which is defined for each individual cast member and scene of the movie. Then, the face, which is produced by facial expression synthesis and re-lighting, is rendered into the face region in the scene image. Therefore animation in which the CG character speaks and expresses his or her feelings as if it were the visitor itself can be generated.

2 Related Works

Face Geometry Acquisition. A variety of 3D range scanners [1] are already commercially available. However, it is difficult to capture the 3D geometry of areas which absorb lasers or sine-wave patterns such as the eyes, the inside of the mouth, and hair. In addition, these scanners are too expensive to use for our work. So we propose an original face scanning system which captures and reconstructs range data in a very short time maintaining high degree of accuracy.

Face Recognition. Wiskott et al developed elastic bunch graph matching using Gabor wavelets [2]. Kawade and Lee also developed techniques to improve this method [3],[4], and have been able to perform facial feature point extraction. This method can reliably extract facial features from frontal face images. Li et al have evaluated this technique using their facial database and, as a result, 77% of automatically extracted feature points were within 5 pixels of the ground truth. Baback and Lian have developed age and gender estimation techniques [5],[6] that are able to estimate age and gender with more than 80% accuracy. FCS system introduces 89 feature points extraction from frontal face image based on [2] with 94% accuracy of face modeling. In FCS, estimated age information is used for casting order and gender information is used to choose pre-scored voice track.

Facial Animation. Since the earliest work in facial modeling and animation [7], generating realistic faces has been the central goal. A technique to automatically generate realistic avatars has been developed using facial images [8]. The CG face model is generated by adjusting a generic face model to target facial feature points, which are then extracted from the image. There are many researches which can construct facial geometry directly from images using morphable facial models [9]-[11]. In FCS system, by fitting generic model onto one frontal face image and then get a model depth from range scanning data at the same time to simplify total face modeling process within 15 seconds now.

Current major facial synthesis techniques include: image-based, photo-realistic speech animation synthesis using morphed visemes [12], a muscle-based face model has been developed [13]-[15], as has a blend shape technique where facial expressions are synthesized using linear combinations of several basic sets of pre-created faces for blending [11],[16]-[19]. Additionally, a facial expression cloning technique has also been developed [20].

The blend shape-based approach applicable to individuals is introduced in FCS to reduce the calculation cost of real-time process and also the labor cost of production process. Basic shape patterns are prepared in generic face model to represent the target facial expression in advance. After fitting process, personalized basic shapes are automatically created and prepared for blend shape process in the story.

Face Relighting. To achieve photorealistic face rendering, human skin's Bi-directional Reflectance Distribution Function (BRDF) and subsurface scattering of light by human skin need to be measured for a personal skin reflectance model. A uniform analytic BRDF is measured from photographs[10]. A Bi-directional Sub-surface Scattering Reflectance Distribution Function (BSSRDF) is developed with parameters based on biophysically-based spectral models of human skin[21]. The method which generates a photorealistic re-lightable 3D facial model is proposed by mapping image-based skin reflectance features onto the scanned 3D geometry of a face[22]. An image-based model, an analytic surface BRDF, and an image-space approximation for subsurface scattering are combined to create highly-realistic facial models for the movie industry[23]. A variant based on this method for real-time rendering on recent graphics hardware is proposed[24]. 3D facial geometry, skin reflectance characteristics and subsurface scattering are measured using custom-built devices[25].

To render faces in FCS, only a face texture (diffuse texture) captured by a range scanner can be used. Furthermore, in "Grand Odyssey" the movie title specially arranged for FCS, there are scenes where a maximum of five characters appear simultaneously. Thus, the rendering time for each character's face is restricted. An important role of our face renderer is to generate images that reduce the differences between visitors' faces and the pre-rendered story images. For skin rendering, first the texture of the scanned face together is used with uniformly specular lighting as a diffuse albedo. Then a hand-generated specular map is used on which speculars on the forehead, cheeks, and tip of the nose become more non-uniformly pronounced than on other parts of the face. Thereby efficiently an image similar in quality to that of the pre-rendered background CG image can be generated.

Immersive Entertainment. The last decade has witnessed a growing interest in employing immersive interfaces, as well as technologies such as augmented/virtual reality (AR/VR) to achieve a more immersive and interactive theater experience. For example, several significant academic research have focused on developing research prototypes for interactive drama [26]-[28]. One technical feature of the current immersive dramatic experience systems is characterized by a reliance on VR/AR devices and environments that create the immersive experience. Typically, players need to wear video-see-through head-mounted displays (HMD) to enter the virtual world.

DIM system is world first an immersive entertainment system commercially implemented as FCS system in Aichi Expo.2005. In DIM movie is generated by replacing the faces of the original roles in the pre-created background movie with audience's own high-realism 3D CG faces. DIM movie is in some sense a hybrid entertainment form, somewhere between a game and storytelling, and its goal is to enable audience easily to participate in a movie as its roles and enjoy an immersive, first-person theater experience.

3 Summary of FCS

Future Cast System (FCS) has two key features: first, it can automatically create a CG character in a few minutes from capturing the facial information of a participant and generating her/his corresponding CG face, to inserting the CG face into the movie; second, FCS makes it possible for multiple people easily to take part in a movie at the same time in different roles, such as a family and a circle of friends. This study is not limited to academic research; 1,630,000 people enjoyed the DIM Movie experience at the Mitsui-Toshiba pavilion at the 2005 World Exposition in Aichi, Japan.

And now it is extended to new entertainment event at HUIS TEN BOSCH, Nagasaki, Japan from March 2007 and also becoming to the most popular attraction there. In this HUIS TEN BOSCH version, the face capturing process is performed only less than 15 seconds. So in case modeling failure is detected by attendant, it is possible to capture face again and again. Fig. 1 shows the processing flow of bringing audience into the pre-created movie.

Processing flow of FCS is as follows.

1. Under the attendant's guidance, participants put their faces to a 3D range-scanner.
2. The captured facial images are transmitted to Face Modeling PCs.
3. The frontal face images are transferred from Face Modeling PCs to Storage Server.
4. 2D thumbnail images are created and sent to Attendant PC to check to see if the captured images are suitable to undergo modeling. If OK, the decision on face modeling is sent back to Face Modeling PCs.
5. In Face Modeling PCs, 3D facial geometry is first generated which make the personal CG face with texture. Personal key shapes are generated for blend shape.
6. The personal face model, age and gender information used for assigning movie roles and the personal key shapes are transmitted to Storage Server.
7. Thumbnail images of participants' face model and information about age and gender is transmitted to Attendant PC from Storage Server. The attendant judges if the participant's face model is suitable to be appeared on the movie. If Yes, Attendant PC automatically assigns a movie role for each participant based on the gender and age estimation. When an attendant found the modeling error happened, it goes back to the process No.1 to capture the guest face again with careful instruction.
8. The decisions on assigning roles to participants are transmitted to Storage Server. A voice track is selected based on the gender information.
9. The information required for rendering movie, including casting information, participant's CG face model, frontal face texture, and customized key shapes for facial expression synthesis, are transmitted to the Rendering Cluster System from Storage Server.

10. Finally, Rendering Cluster System embeds participant's face into the pre-created background movie images using the received information and the completed movie images are projected onto a screen.

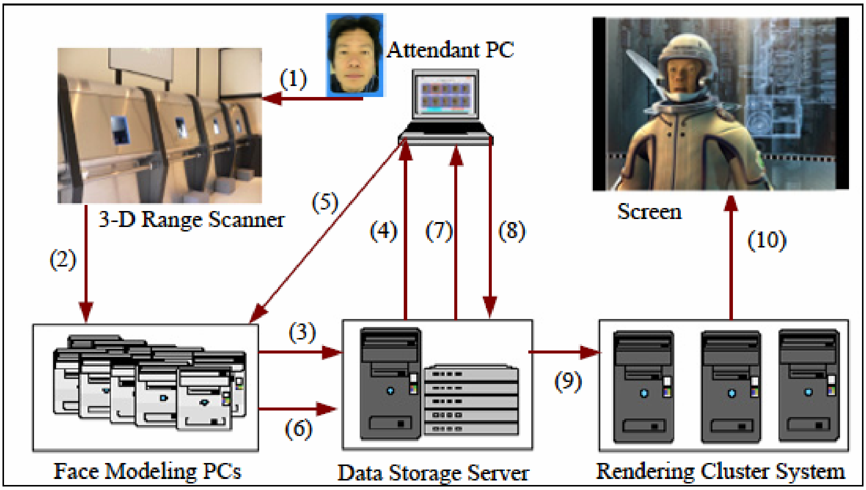


Fig. 1. Processing Flow of FCS

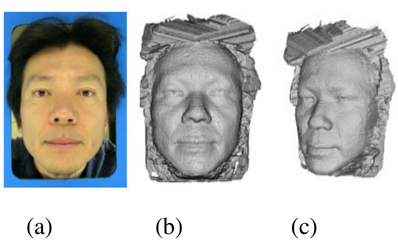


Fig. 2. Face Modeling Result. (a) frontal image. (b) (c) constructed geometry.

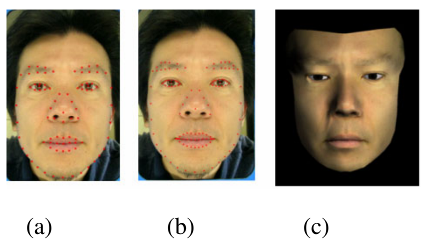


Fig. 3. Personal Face Model. (a) 89 feature points. (b) 128 feature points. (c) face model with personality.

Face Capturing System. A non-contact, robust and inexpensive 3D range-scanner is developed to generate participant facial geometry and textural information. This scanner consists of a few CCD digital cameras and two slide projectors which are able to project stripe patterns. They are placed in a half circle and can instantly capture face images in few seconds. The central camera captures the frontal face image to generate texture mapping (Fig. 2.a). The hybrid method with an Active-Stereo technique and a Shape-from-Silhouette technique [1] are used to reconstruct 3D geometry. Fig. 2 shows a frontal face image and its corresponding 3D face geometry constructed by a scanner system.

Face Modeling. Face modeling is completed in approximately 2 minutes in Aichi Expo. version, 15 seconds in HUIS TEN BOSCH version of FCS system and now up-to-date scanner can complete it less than 10 seconds with the highest accuracy. Overall this process includes four aspects: facial feature point extraction, face fitting, face normalization and gender and age estimation. Currently, we proposed “Active Snap Shot” which can generate a texture mapped personal head model in about 2 seconds only with 1 single 2D snap shot and 2mm vertex accuracy.

Face Fitting. In the face fitting phase, a generic face model is deformed using the scattered data interpolation method (Radial Basis Functions transformation, [32]-[34]) which employs extracted feature points as targets. This technique enables us to deform the model smoothly. However, we found that deformed vertices on the model do not always correspond to the outlines of facial areas other than the target points. According to our own experience, it is important that the vertex of the model conform closely to the outline of the eyes and the lip-contact line on the facial image because these locations tend to move a lot when the model performs an action such as speaking or blinking. We therefore must accurately adjust the model for the feature points corresponding to the outline of the eyes, lips, and especially, the lip-contact line. Consequently, we interpolate these areas among the 89 feature points (Fig.3.a) using a cubic spline curve and then resample 128 points (Fig.3.b) on an approximated curve to achieve correspondence between the feature points. This process is performed for eyes, eyebrows, and the outline of the lip.

Real-time Facial Expression Synthesis. We selected the blend shape approach to synthesize facial expression, since it can operate in real-time and it’s easy to prepare suitable basic expression patterns with which animators can produce and edit specific facial expressions in each movie scene. Also, the parameters for facial expression are very simple because they are only expressed with the blending ratio of each basic face. We first create universal 34 facial expression key-shapes based on generic face mesh, and then calculate the displacement vectors between the neutral mesh (generic face mesh) and the specific expressions’ mesh (one of 34 key shapes) such as mouth opening. In the current DIM system, basic face is constructed by physics based facial muscle control adjusted to each audience’s estimated facial muscle structure.

Relighting Face. The goal of relighting faces is to generate a matching image between the participant’s face and the pre-created movie frames. Considering the requirement for real-time rendering, we implemented a pseudo-specular map for creating facial texture. Specifically, we chose to use only the reflectional characteristic of participants’ skin which is acquired from the scanned frontal face images; however, the images still contain a combination of ambient, diffuse and specular refection. We therefore devised a lighting method to ensure a uniform specular reflection for the whole face by installing an alignment of florescent lights fitted with white diffusion filters. To represent skin gloss, we developed a hand-generated specular map on which speculars on the forehead, cheeks, and tip of the nose become more non-uniformly pronounced than other parts of the face. The intensity and sharpness of the speculars were generated by white Gaussian noise. The map can be applied to all participants’ faces. Additionally, we approximated subsurface scattering and hair lighting effects by varying intensity on the outlines of the characters’ faces.

Currently, personal specular feature is customized to each individual by the captured data from custom made skin sensor.

Automatic Casting and Voice Selection. FCS can automatically select a role from the background movie for each participant according to age and gender estimation. In addition, we pre-recorded a male and female voice track for each character, and voices were automatically assigned to characters according to the gender estimation results.

In current DIM, an individualized voice is generated by huge voice actor database.

Real-time Movie Rendering. The rendering cluster system consists of 8 PCs tasked for image generation of movie frames (Image Generator PC, IGPC) and a PC for sending the image to a projector (Projection PC, PPC). The PPC first receives the control commands, time codes, and participants' CG character information from Data Storage Server, and then sends them to each IGPC. The IGPCs render the movie frame images using a time division method and send them back to the PPC.

The performance of graphics hardware is advanced drastically, current DIM system can generate and render multiple characters with whole body motion and rendered hair as well as facial expression on real-time process.

Evaluation. FCS has been successfully exhibited from 3/25 to 9/25 at the 2005 World Expo. in Aichi, Japan, and showed a high success rate (94%) of bringing audience into the movie. However, evaluation result shows self-awareness rate is only 65%. Now up-to-date DIM movie can make self-awareness rate to 86% by introducing new technology described next.

4 Up-to-Date DIM Movie

In the former FCS system, 3D geometry and texture of audience face without glasses can be reproduced precisely and instantly in a character model. However, facial expression is controlled only by a blend shape animation, so personality of smile is not reflected in animation at all. Also the head of character is covered by a helmet in the story "Grand Odyssey" because of the huge calculation cost of hair animation and body motion, which are treated as a back ground image with fixed size and fixed costume assigned to the casting. So sometimes a few of audience cannot identify himself on the screen even if his family or friends can identify him. The progress in GPU performance is so high that real-time rendering and motion control of full body and hair become possible now.

Head Modeling. To emphasize personality feature on character facial animation, a head and hair modeling are inevitable. After scanning face, we tried to fit generic skin head model to captured face geometry and then a pre-designed wisp model of hair is generated automatically on this head model. To reduce the calculation cost, each wisp of hair is modeled by a flat transparent sheet with hair texture and an extended cloth simulation algorithm is applied to make hair animation.

Figure 4 shows the generated hair variation from an automatically constructed head model. The technical detail is described in [29]. To change hair style makes each character more attractive and impressive.

Fig. 4. Head Model with Hair Style Variation

Face Modeling. Blend shape facial animation is so simple that it is very convenient for creator to direct a specific expression and it takes very low CPU cost to synthesize on real-time process. However, to reflect personal characteristics on facial animation, physics based face model is effective because it is easy to give variation in motion by changing the location and stiffness of each facial muscle.

In the most recent DIM system, after fitting a face wireframe to scanned face range data, the location of each standard muscle can be decided automatically. By modifying the location of muscles, a variety of personal individuality can be generated in a same face model. This optimization of the location and physical character of each muscle is performed based on a captured face video appearing smile without any landmarks on face. The detail is expressed in [29].

Individual skin feature is captured by custom made skin sensor and reflected to the rendering of personalized characters [32].

Body and Gait Modeling. Character model with a variety of body size is pre-constructed and optimum size is decided according to the captured silhouette image. By fitting the size of neck, head, face and hair to the body size, a full body character model is assembled by morphing of these basic body shapes. Also personality in gait is modeled automatically by combination of key motions according to the silhouette estimation captured by 2 cameras. This technique is expressed in [30].

Voice Modeling. In former FCS system, a prerecorded voice of an actor or actress is used as a substitute for that of each participant. The substitute voice is selected by only each participant's gender information which is estimated based on the scanned face shape without consideration of other information such as age and voice quality. This caused some sense of discomfort for those who perceive the voice of the character to be different from their own or the people they know. Therefore we tried to choose or synthesize voice with personality based on the similarity measurement [31].

Immersive Experience with First Person Perspective. DIM project also produces both visual and audible immersive environment with first person perspectives. Prototype for high fidelity 3D audio recording and play-back [34] and omnidirectional panoramic video recording and screen system [33] have been constructed.

5 Conclusions

In this paper, an innovative visual entertainment system DIM is presented. The face modeling process, the core technology of our system, can automatically generate visitors' CG faces with an accuracy of 93.5%. So the former Future Cast System provides audiences with a movie in which only their faces are embedded.

Fig. 5. Personalized Character In "GrandOdyssey"

In current DIM system, how to improve the audience's sense of personal identification while watching the movie is considered. The key lies in generating CG characters which share the visitors' physical characteristics, such as hairstyle, body shape, gait motion, and voice personality. By introducing these factors, self-awareness rate is improved to 86% from 65% of former FCS system.

The detail of DIM project is as follows.

http://www.mlab.phys.waseda.ac.jp/DIM/home_e.html

References

1. Fujimura, K., et al.: Multi-camera 3d modeling system to digitize human head and body. In: Proceedings of SPIE, pp. 40–47 (2001)
2. Wiskott, L., et al.: Face recognition and gender determination. In: FG 1999, pp. 92–97 (1999)
3. Kawade, M.: Vision-based face understanding technologies and applications. In: Proc. of 2002 International Symposium on Micromechatronics and Human Science, pp. 27–32 (2002)
4. Zhang, L., et al.: A fast and robust automatic face alignment system. In: ICCV 2005 (2005)
5. Moghaddam, B., Yang, M.: Gender classification with support vector machines. In: Proceedings of the fourth IEEE FG 2000, pp. 306–311 (2000)
6. Lian, H., et al.: Gender recognition using a min–max modular support vector machine. In: International Conference on Natural Computation, pp. 438–441 (2005)
7. Parke, F.I.: Computer generated animation of faces. In: ACM 1972: Proceedings of the ACM annual Conference (1972)
8. Lyons, M., et al.: Avatar creation using automatic face processing. In: Proceedings of the sixth ACM international conference on Multimedia, pp. 427–434 (1998)
9. Pighin, F., et al.: Synthesizing realistic facial expressions from photographs. In: SIGGRAPH 1998 (1998)
10. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d faces. In: Proceedings of SIGGRAPH 1999, pp. 187–194 (1999)
11. Blanz, V., et al.: Reanimating faces in images and video. In: Computer Graphics Forum 2003 (2003)

12. Ezzat, T., et al.: Trainable videorealistic speech animation. In: SIGGRAPH 2002, pp. 388–398 (2002)
13. Lee, Y., Terzopulus, D., Waters, K.: Realistic modeling for facial animation. In: Proceedings of SIGGRAPH 1995, pp. 55–62 (1995)
14. Choe, B., Lee, H., Ko, H.: Performance-driven muscle-based facial animation. In: Proceedings of Computer Animation, pp. 67–79 (2001)
15. Sifakis, E., Neverov, I., Fedkiw, R.: Automatic determination of facial muscle activations from sparse motion capture marker data, pp. 417–425 (2005)
16. DeCarlo, D., Metaxas, D.: The integration of optical flow and deformable models with applications to human face shape and motion estimation. In: CVPR 1996, pp. 231–238 (1996)
17. Joshi, P., Tien, W.C., Desbrun, M., Pighin, F.: Learning controls for blend shape based realistic facial animation. In: SCA 2003, pp. 187–192 (2003)
18. Xiang, J., Chai, Xiao, J., Hodgins, J.: Vision-based control of 3d facial animation. In: SCA 2003, pp. 193–206 (2003)
19. Vlasic, D., Brand, M., Pfister, H., Popovi, J.: Face transfer with multilinear models. ACM Transactions on Graphics 24(3), 426–433 (2005)
20. Noh, J.Y., Neumann, U.: Expression cloning. In: SIGGRAPH 2001, pp. 277–288 (2001)
21. Krishnaswamy, A., Baranoski, G.V.G.: A biophysically-based spectral model of light interaction with human skin. Comput. Graph. Forum, 331–340 (2004)
22. Debevec, P., Hawkins, T., Tchou, C., Duiker, H.P., Sarokin, W., Sagar, M.: Acquiring the reflectance field of a human face. In: SIGGRAPH 2000, pp. 145–156 (2000)
23. Borshukov, G., Lewis, J.P.: Realistic human face rendering for ”the matrix reloaded. In: ACM SIGGRAPH 2003 Technical Sketch (2003)
24. Sander, P.V., Gosselin, D., Mitchell, J.L.: Real-time skin rendering on graphics hardware. In: SIGGRAPH 2004, Sketches, p. 148 (2004)
25. Weyrich, T., et al.: Analysis of human faces using a measurement-based skin reflectance model. In: SIGGRAPH 2006, pp. 1013–1024 (2006)
26. Cavazza, M., Charles, F., Mead, S.J.: Interacting with virtual characters in interactive storytelling. In: Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems: part 1, pp. 318–325 (2002)
27. Cheok, A.D., Weihua, W., Yang, X., Prince, S., Wan, F.S., Billinghurst, I.M., Kato, H.: Interactive theatre experience in embodied + wearable mixed reality space. In: Proc. of ISMAR 2002, pp. 59–317 (2002)
28. Mateas, M., Stern, A.: Facade: An experiment in building a fully-realized interactive drama. In: Proceedings of Game Developers' Conference: Game Design Track (2003)
29. Maejima, A., et al.: Realistic Facial Animation by Automatic Individual Head Modeling and Facial Muscle Adjustment. In: HCI International 2011, ID:1669 (to appera, 2011)
30. Makihara, Y., et al.: The Online Gait Measurement for Characteristic Gait Animation Synthesis. In: Proc.of HCI International 2011, ID:1768 (to appear, 2011)
31. Shin-ichi, K., et al.: Personalized Voice Assignment Techniques for Synchronized Scenario Speech. In: HCI International 2011, ID:2218 (to appear, 2011)
32. Mashita, T., et al.: 'Measuring and Modeling of Multi-layered Subsurface Scattering for Human Skin. In: Proc.of HCI International 2011, ID:2219 (to appear, 2011)
33. Kondo, K., et al.: Providing Immersive Virtual Experience with First-person Perspective Omnidirectional Movies and Three Dimensional Sound Field. In: HCI International 2011, ID:1733 (to appear, 2011)
34. Enomoto, S., et al.: 3-D sound reproduction system for immersive environments based on the boundary surface control principle. In: Proc. HCI International 2011, ID: 1705 (to appear, 2011)

A Realtime and Direct-Touch Interaction System for the 3D Cultural Artifact Exhibition

Wataru Wakita[1], Katsuhito Akahane[2], Masaharu Isshiki[3], and Hiromi T. Tanaka[1]

[1] Department of Human and Computer Intelligence, College of Information Science and Engineering, Ritsumeikan University, Kusatsu, Japan
{wakita,hiromi}@cv.ci.ritsumei.ac.jp
[2] Precision and Intelligence Lab, Tokyo Institute of Technology, Yokohama, Japan
kakahane@hi.pi.titech.ac.jp
[3] Department of Electrical and Electronic Engineering and Computer Science, Ehime University, Matsuyama, Japan
isshiki@cs.ehime-u.ac.jp

Abstract. We propose a realtime and direct-touch interaction system for 3D cultural artifact exhibition based on a texture-based haptic rendering technique. In the field of digital archive, it is important to archive and exhibit the cultural artifact at the high-definition. To archive the shape, color and texture of the cultural artifact, it is important to archive and represent not only visual effect but haptic impression. Therefore, multimodal digital archiving, realtime multisensory rendering, and intuitive and immersive exhibition system are necessary. Therefore, we develop a realtime and direct-touch interaction system for the 3D cultural artifact exhibition based on a texture-based haptic rendering technique. In our system, the viewer can directly touch a stereoscopic vision of 3D digital archived cultural artifact with the string-based and scalable haptic interface device "SPIDAR" and vibration motor.

Keywords: Digital Museum, Virtual Reality, Computer Graphics, Haptics.

1 Introduction

We are working on a digital museum project of the "Gion Festival in Kyoto" [1], [2]. The floats in the "Gion Festival in Kyoto" are decorated with various accessories. Therefore, the floats are described as a "Moving Museum". In this digital museum project, in particular, we are working on a digital archiving of the "Gion Festival in Kyoto" by multisensory information such as the visual sense and the haptic sense.

In the field of digital archive, it is important to archive and exhibit the cultural artifact at the high-definition. To archive the shape, color and texture of the cultural artifact, it is important to archive and represent not only visual effect but haptic impression. To exhibit the digital archived 3D model naturally, multisensory digital archiving and interactive, intuitive and immersive exhibition system is necessary.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 197–205, 2011.

Moreover, to reduce the graphic and haptic rendering cost, a realtime rendering is necessary. Generally, to achieve the realtime rendering in graphic process, at least 30-60 Hz update rate is necessary. In haptic process, to represent the soft objects, at least 300 Hz update rate is necessary, and to represent the hard objects, at least 10 kHz update rate is necessary while considering the graphic rendering cost.

Therefore, we develop a realtime and direct-touch interaction system for the 3D cultural artifact exhibition with the string-based and scalable haptic interface device "SPIDAR" [3] and vibration motors.

2 Related Work

Recently, various haptic rendering devices have been developed, and various haptic rendering techniques to touch the virtual object have been proposed. The penalty-based haptic rendering method [4], [5] is a basic approaches to represent the polygon wall, has several problems such as passing through, discontinuous force and vibration. To solve these problems, Zilles *et al.* proposed a constraints-based God-object method [6]. However, their method has the same problems such as passing through, discontinuous force and vibration in haptic rendering for the high-definition virtual object. On the other hand, several texture-based haptic rendering techniques have proposed to represent the asperity of the interior of the polygon according to the 2D image [7], [8], [9], [10]. In our previous work, we proposed a texture-based haptic rendering technique for the pseudo-roughness on the surface of the low-polygon virtual object using height map and normal map, and we developed a material system under haptic rendering for pseudo-roughness on the low-polygon object surface [9]. In this system, a difference of the haptic impression is represented by changing magnitude and/or direction of the reaction force dynamically according to the pixel value of the object surface which mapped the haptic texture which converted surface height, stiffness and friction into the 2D image. Moreover, we have proposed a realtime haptic rendering technique for representation of the high-definition model surface using the low-polygon model, distance map and normal map [10]. In this approach, the reaction force is calculated according to the pixel value of the low polygon model surface which mapped the haptic texture which converted the geometric difference of the high-polygon model and the low-polygon model into the 2D image. However, these techniques are not based on the measurement. To represent the high-definition virtual object, it is necessary to model the virtual object based on the measurement. The same can be said for digital archive. Therefore, in our previous work, we captured the surface structure of the materials in the real world with OGM (Optical Gyro Measuring Machine) and generated the normal map [11] which has surface asperity information. Then we modeled the tactile sense by vibration signals based on normal map, and we developed a haptic rendering system for a 3D noh-cloth model based on the measurement with the string-based haptic interface device SPIDAR and vibration speakers [12].

2.1 Multisensory System of the Cultural Heritage

As related work of the cultural heritage, Christou *et al.* [13] proposed a versatile large-scale multimodal VR system for cultural heritage visualization. Their exhibition system is the Cave-like multimodal system, and their system enables haptic interaction with two haptic arm. On the other hand, Carrozzino *et al.* [14] proposed a large-scale multimodal and immersive VR system. Their exhibition system is the Cave-like multimodal system, and their system enables haptic interaction with exoskeleton haptic arm. These works enables to represent the shape, color, and texture of the cultural artifact with the haptic device. However, these systems are impossible to directly touch to the digital archived model. To exhibit the digital archived 3D model naturally, a direct touchable immersive exhibition system is necessary.

2.2 Direct-Touchable Multisensory System

As related work of the direct-touchable multisensory system, Inami *et al.* [15] and Arsenault *et al.* [16], [17] proposed a system with half mirror and haptic arm device. Vallino *et al.* [18], Bianchi *et al.* [19], and Sandor *et al.* [20], [21] proposed a AR/MR system with HMD and haptic arm device. Brederson *et al.* [22] and Ikits *et al.* [23], [24] proposed a system with 3D projector and haptic arm device. These system enables direct-touch for the graphic models with the haptic arm device. However, it is difficult to touch the large-scale cultural artifact because there is a limitation over the range of arm movement. On the other hand, Yoshida et al. [25] proposed RePro3D with vibration motor and retro-reflective projection technology. However, it is impossible to represent the kinematic sense.

To exhibit the cultural artifact naturally, a direct-touchable exhibition interface corresponding to the size of the cultural artifacts is necessary (see Figure 1). Therefore, in our work, we use the string-based and scalable haptic interface device "SPIDAR" [3] and vibration motors.

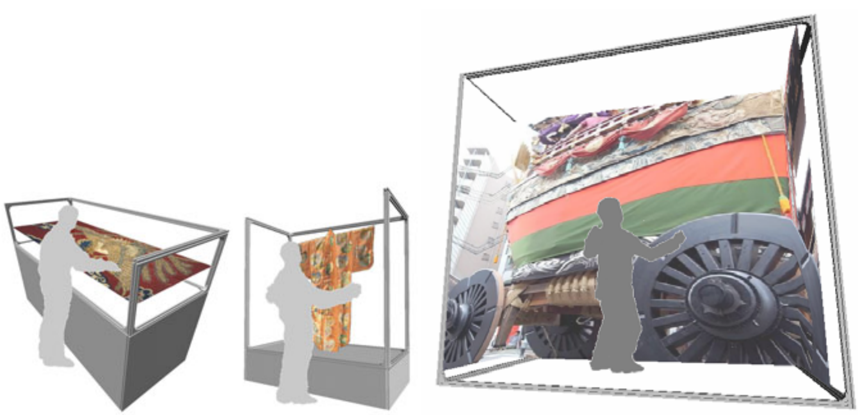


Fig. 1. Direct-touchable multisensory exhibition interface corresponding to the size of the cultural artifacts with scalable haptic interface device SPIDAR

3 Direct-Touchable Multisensory Exhibition System

Figure 2 shows a direct-touch able multisensory exhibition system with SPIDAR. SPIDAR is on top of a screen. SPIDAR has ability to control the 3DOF position and to present the 3DOF forces. A grip part is attached to 4 strings from 4 motors with an encoder. The strings length got from each encoder's data is used to measure the grip's position. The strings tension from each motor is displayed the feedback forces.

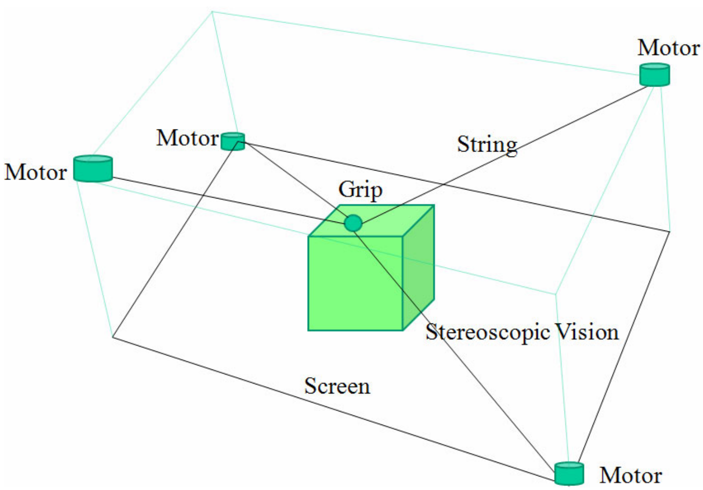


Fig. 2. Direct-touchable multisensory exhibition system with SPIDAR

Users attach the vibration motors on fingers, users can feel kinematic sense and tactile sense moving the fingar tip and tracing the stereoscopic 3D virtual objects on the screen. Reaction force and vibration signal is calculated with the finger's move direction vector and surface asperity of collision point. A vibration signal is converted to voltage by D/A converter.

3.1 Direct-Touch Interaction for the Stereoscopic Vision

To touch the stereoscopic objects naturally with a SPIDAR, a grip position in the device space is required to match the virtual space according to the camera space.

To represent the 3D shape with SPIDAR, firstly, the intersection is detected between the finger tip and 3D model surface on the screen. Secondly, if they are crossed in the intersection detection and have the possibility of contact, the polygon height is changed according to the pixel value of the height map in relation to intersection point and the polygon is replicated [9], [10] (see Figure 3).

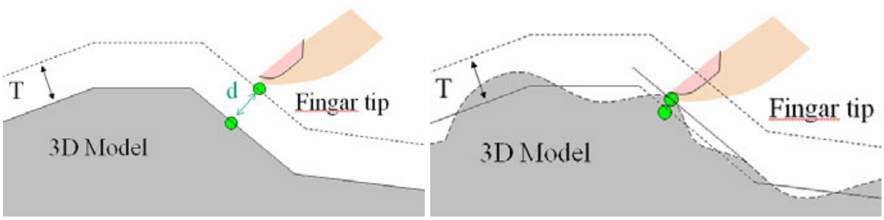


Fig. 3. Direct-touch interaction for the 3D model based on our texture-based haptc rendering technique

Finally, the intersection is detected again between the finger tip and a copy polygon, and the reaction force is calculated according to the penetration depth and the pixel value of the friction map and stiffness map [10]. Our haptic rendering technique is based on a constraint-based God-object method [6].

In our system, we display the vibration signal transform to voltage and input to vibration motors. The signal is calculated by the inner product finger tracing direction and the normal vector in the contact point of normal map.

4 Digital Archive

4.1 Graphic Modeling

We used the laser range scanner "VIVID" for the measurement of the shape, and we used a high-resolution multiband imaging camera for measurement of the color and spectral reflectance.

Figure 4, 5 shows measurement data of the woven cultural artifact "Hirashaji Houou Monyou Shishu" of "Fune-hoko" of "Gion Festival in Kyoto". Figure 4 shows a height image data (height map) that was generated from measured range data by the laser range scanner, and Figure 5 shows a color image data (color map) by the multiband camera.

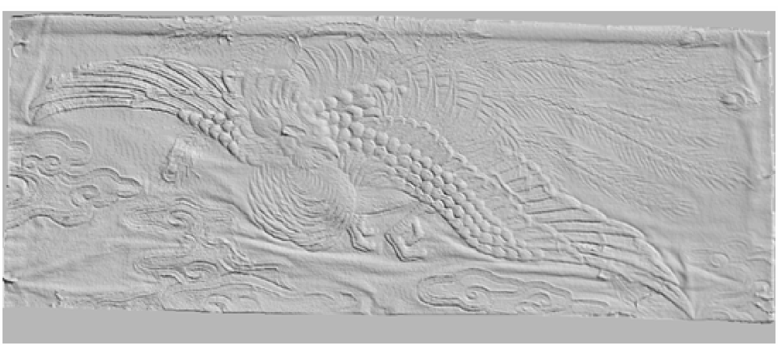


Fig. 4. Measured height image data of the "Hirashaji Houou Monyou Shishu"

Fig. 5. Measured color image data of the "Hirashaji Houou Monyou Shishu"

The measured range data have 612,522 (1193x512) vertices, and measured color image data have 73,300,500 (13650x5370) pixels. To reduce the graphic rendering cost, we reduced a 3D polygon model and we mapped a measured 2D color map to a reduced 3D polygon model (see Figure 6).

Fig. 6. 3D Digital Archived Model of the "Hirashaji Houou Monyou Shishu"

4.2 Haptic Modeling

To reduce the haptic rendering cost, we used our texture-based haptic modeling and rendering technique [26]. Firstly, we created a normal map from height map (see Figure 7).

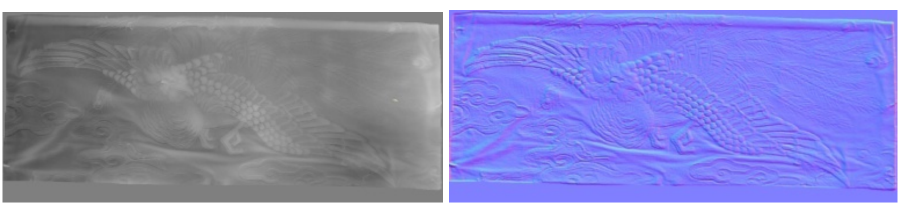


Fig. 7. Height map (*left*) and Normal map(*right*) of the "Hirashaji Houou Monyou Shishu"

This normal map is used to represent the surface gradient, where the RGB values correspond to the XYZ coordinates of the normal vector. The height map is used to represent the surface height, where the surface height is changed according to the grayscale value (white is high and black is low elevation). We mapped these maps to low-polygon model which is used for the haptic rendering, and is not used for the graphic rendering.

5 Results

We developed a direct-touch interaction system for the digital archived 3D cultural artifact exhibition (see Figure 8). Our system is composed of a display system and an 3D application. A display system consists of a graphic part and haptic part. In graphic part, we used a rear projector screen (1000x750 mm) and the stereoscopic projector "DepthQ HD". The stereoscopic vision is projected to the bottom projector screen with a mirror. A haptic part is on top of a projector screen, and at one with a graphic part. In haptic part, we used SPIDAR and vibration motors. We used two 2.33 GHz Intel(R) Xeon(R) CPU E5410, NVIDIA Quadro FX 580 graphics card with 512MB video memory, 16GB RAM, Windows VISTA 64bit, and NVIDIA 3D Vision. The graphic process is 120 Hz update rate, and haptic process is 1 kHz update rate.

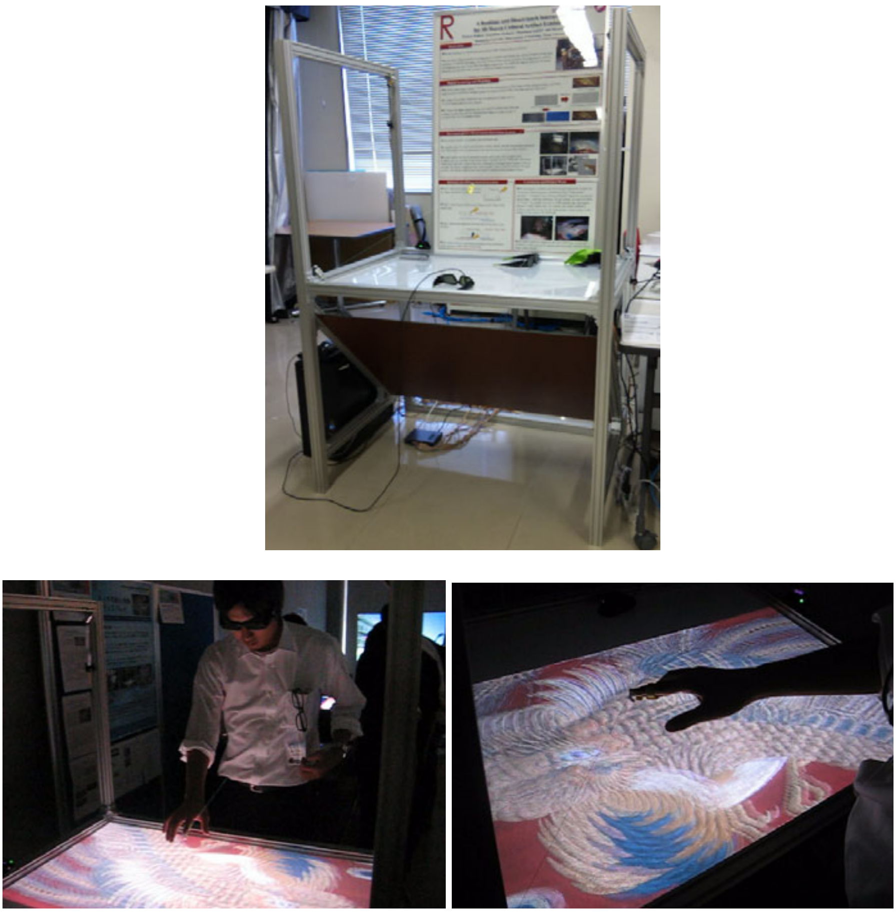


Fig. 8. Realtime and direct-touch interaction system for the digital archived 3D cultural artifact exhibition

6 Conclusion and Future Work

We developed a realtime and direct-touch interaction system for the 3D cultural artifact exhibition with the string-based and scalable haptic interface device "SPIDAR" and vibration motors. Specifically, firstly we archived the cultural artifact "Tenmizuhiki" tapestries "Hirashaji Houou Monyou Shishu" of "Fune-hoko" of "Gion Festival in Kyoto". Secondly, we developed a exhibition system with the stereoscopic projector, SPIDAR and vibration motors based on our texture-based technique. However, in our system, the stiffness properties are not based on the measurement data. Therefore, we plan to measure various materials in the real world.

Acknowledgments. This research has been conducted partly by the support of the Digital Museum Project in the Ministry of Education, Sports, Science and Technology, Japan. We would like to thank the Gion-Matsuri Fune-hoko Preservation Society, a generous collaborator of this project.

References

1. Tanaka, H.T., Yano, K., Hachimura, K., Nishiura, T., Choi, W., Fukumori, T., Furukawa, K., Wakita, W., Tsuchida, M., Saiwaki, N.: Digital Archiving of the World Intangible Cultural heritage “Gion Festival in Kyoto”: Reproduction of “Fune-boko” Float of the Gion Festival Parade in “Virtual Kyoto”. In: Proceedings of the ASIAGRAPH 2010, Shanghai, China (2010)
2. Tanaka, H.T., Hachimura, K., Yano, K., Tanaka, S., Furukawa, K., Nishiura, T., Tsutida, M., Choi, W., Wakita, W.: Multimodal Digital Archiving and Reproduction of the World Cultural Heritage “Gion Festival in Kyoto”. In: Proceedings of the ACM SIGGRAPH Sponsored 9th International Conference on VR Continuum and Its Applications in Industry (VRCAI 2010), COEX, Seoul, Korea (2010)
3. Sato, M.: Development of string-based force display: Spidar. In: Proceedings of the 8th International Conference on Virtual Systems and Multi Media (VSMM 2002), Gyeongju, Korea (2002)
4. Mark, W.R., Randolph, S.C., Finch, M., Verth, J.M.V., Taylor II, R.M.: Adding Force Feedback to Graphics Systems: Issues and Solutions. In: Proceedings of the SIGGRAPH 1996, New Orleans, U.S.A, pp. 447–452 (1996)
5. Massie, T.H., Salisburg, J.K.: The PHANTOM Haptic Interface: A Device for Probing Virtual Objects. In: Proceedings of the ASME Winter Annual Meeting, Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, DSC, Chicago, U.S.A, vol. 55(1), pp. 295–301 (1994)
6. Zilles, C., Salisbury, K.: A constraint-based God-object method for haptic display. In: Proceedings of the IEEE/RSJ International Conference Intelligent Robots and Systems, Pittsburgh, U.S.A, pp. 146–151 (1995)
7. Stanney, K.: Haptic Rendering In Virtual Environments. Handbook of Virtual Environments. In: Basdogan, C., Srinivasan, M.A. (eds.), pp. 117–134. Lawrence Erlbaum Associates, Inc., London (2001)
8. Theoktisto, V., Fairén, M., Navazo, I., Monclús, E.: Rendering detailed haptic textures. In: Proceeding of the 2nd Workshop in Virtual Reality Interactions and Physical Simulations (VRIPHYS 2005), Pisa, Italy, pp. 16–23 (2005)

9. Wakita, W., Ido, S.: A Material System under Haptic Rendering for Pseudo-Roughness. IEICE Trans. Inf. & Syst (Japanese Edition) J91-D(8), 2061–2070 (2008)
10. Wakita, W., Ido, S.: A Haptic Rendering for High Polygon Model Using Distance Map and Normal Map. IPSJ Journal 49(7), 2509–2517 (2008)
11. Ozaki, R., Nishiwaki, Y., Takeda, Y.: Weave pattern modeling of silk-like fabrics from multi-illuminated HDR image analysis. Transactions of the Virtual Reality Society of Japan 14, 315–324 (2009)
12. Nomura, K., Yin, X., Sakaguchi, Y., Tanaka, H.: Modeling system of tactile feeling based on photometoric images analysis. IEICE Trans. Inf. & Syst (Japanese Edition) J93-D(10), 2326–2329 (2010)
13. Christou, C., Angus, C., Celine Loscos, C., Dettori, A., Roussou, M.: A Versatile Large-Scale Multimodal VR System for Cultural Heritage Visualization. In: Proceeding of the ACM symposium on Virtual Reality Software and Technology (VRST 2006), Limassol, Cyprus, pp. 133–140 (2006)
14. Carrozzino, M., Bergamasco, M.: Beyond virtual museums: Experiencing immersive virtual reality in real museums. Journal of Cultural Heritage 11(4), 452–458 (2010)
15. Inami, M., Kawakami, N., Sekiguchi, D., Yanagida, Y., Maeda, T., Tachi, S.: Visuo-Haptic Display Using Head-Mounted Projector. In: Proceedings of the IEEE Virtual Reality 2000, pp. 233–240 (2000)
16. Arsenault, R., Ware, C.: Eye-hand co-ordination with force feedback. In: Proceedings of the ACM Conferernce on Human Factors in Computing Systems (CHI 2000), pp. 408–414 (2000)
17. Arsenault, R., Ware, C.: The Importance of Stereo and Eye Coupled Perspective for Eye-Hand Coordination in Fish Tank VR. Presence 13(5), 549–559 (2004)
18. Vallino, J., Brown, C.: Haptics in augmented reality. In: Proceedings of the IEEE. Int'l Conf. on Multimedia Computing and Systems (ICMCS), vol. 1, pp. 195–200 (1999)
19. Bianchi, G., Knoerlein, B., Szekely, G., Harders, M.: High precision augmented reality haptics. In: Proceedings of EuroHaptics, pp. 169–178 (2006)
20. Sandor, C., Uchiyama, S., Yamamoto, H.: Visuo-Haptic Systems: Half-Mirrors Considered Harmful. In: Proceedings of the WHC 2007, pp. 292–297 (2007)
21. Sandor, C., Kuroki, T., Uchiyama, S., Yamamoto, H.: Exploring Visuo-Haptic Mixed Reality. Technical Report PRMU 2006-199 IEICE, Tokyo, Japan, vol. 106, pp. 31–36 (2007)
22. Brederson, J.D., Ikits, M., Johnson, C., Hansen, C.: In: Proceedings of the PHANToM Users Group Workshop, Aspen, CO, pp. 46–49 (2000)
23. Ikits, M., Hansen, C., Johnson, C.: A Comprehensive Calibration and Registration Procedure for the Visual Haptic Workbench. In: Proceedings of the Eurographics Workshop on Virtual Environments, Zurich, Switzerland, pp. 247–254 (2003)
24. Ikits, M., Brederson, J.D.: In: Hansen, C., Johnson, C. (eds.) Visualization Handbook, pp. 431–447. Academic Press, London (2004)
25. Yoshida, T., Kamuro, S., Minamizawa, K., Nii, H., Tachi, S.: RePro3D: Full-Parallax 3D Display Using Retro-Reflective Projection Technology. In: Proceedings of the ACM SIGGRAPH 2010 Emerging Technologies, Los Angeles, California (2010)
26. Wakita, W., Murakami, K., Ido, S.: Development of a Texture-Based Haptic Modeling System. IEICE Trans. Inf. & Syst. (Japanese Edition) J91-D(12), 2773–2780 (2008)

Digital Display Case: A Study on the Realization of a Virtual Transportation System for a Museum Collection

Takafumi Watanabe[1], Kenji Inose[1], Makoto Ando[1], Takashi Kajinami[2], Takuji Narumi[2], Tomohiro Tanikawa[2], and Michitaka Hirose[2]

[1] TOPPAN PRINTING CO., LTD. 1-3-3 Suido, Bunkyo-ku, Tokyo, Japan
[2] The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
{takafumi.watanabe,Kenji.Inose,makoto.ando}@toppan.co.jp,
{kaji,narumi,tani,hirose}@cyber.t.u-tokyo.ac.jp

Abstract. This paper describes our proposed virtual transportation system. Our proposed system is a display case for use at art museums, which is based on computer graphics and image-based rendering (IBR) techniques. Using this system, anyone can simply create and realistically represent virtual cultural assets. This system consists of two main components: a display unit and a capture unit. The display unit is in the shape of a conventional display case in order to represent virtual cultural assets. The capture unit, which is created by attaching cameras to a conventional display case, reconstructs cultural assets using IBR techniques. In our experiment, we implemented a basic system using View Morphing as the IBR technique. The results show that this system can represent virtual cultural assets as 3D objects on the display unit by using arbitrary view images that are interpolated by View Morphing.

Keywords: digital display case, digital museum, image based rendering, virtual reality.

1 Introduction

In conventional art museums, visitors are not supplied with sufficient information to understand the background of heritage and cultural assets without learning about them in advance. For example, it is difficult for visitors to understand the background of cultural assets, such as the purpose of use, the excavation site, and historical events that took place at the same time. Therefore, new techniques for teaching background information have been researched in recent years. A "digital museum" is a new concept for teaching the background information of heritage and cultural assets.

As an example of a digital museum, the Tokyo National Museum in Japan and the Palace Museum in China have installed a Virtual Reality Theater to present virtual heritage and cultural assets reconstructed by digital archive technologies [1] [2].

We have been researching on a digital museum project based on augmented reality techniques. Our project investigates effective techniques for use in exhibitions of digital museums. These techniques will give more background information to visitors by overlapping the information with real heritage and cultural assets with the use of

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 206–214, 2011.

augmented reality techniques. In this paper, as one of our projects, we propose a new exhibition system for a display case.

2 Outline of the Digital Display Case

We previously proposed a digital display case [3] in response to the following requirements of an art museum:

1. The external appearance should resemble a display case in order to create a seamless arrangement with the conventional display cases in the gallery, and also to easily engage the visitors with the system.
2. The images displayed on the digital display case must be accurately correlated with the viewer's eye gaze in real time, in order for the viewer to perceive the complete illusion as real, that is, as if there are real objects in the case.
3. To enable the presentation of the internal structure or long-term changes in the cultural assets, rich digital content must be produced.
4. Visitors should manipulate the virtual cultural assets with a force feedback mechanism.

Figure 1 shows the design of the digital display case. It consists of four 3D displays on each side, a manipulating device to manipulate the position of a virtual cultural asset, and 3D glasses with a motion tracking device. The virtual cultural assets in the digital display case are represented as a 3D model that is produced by measuring the shape using a 3D scanner. The motion tracking device measures the position of the viewer's eyes, and the view on the digital display case is changed in accordance with the viewer's position. Using the digital display case, the visitors can experience the same views as with cultural assets in conventional display cases, because it appears that the virtual cultural assets have been placed in the display case. Furthermore, visitors can move and rotate objects by using the manipulating device. The digital display case can easily display various cultural assets by switching the 3D objects at the user's discretion.

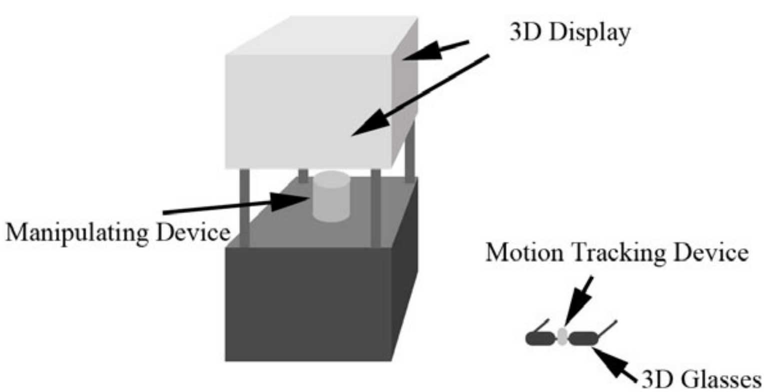


Fig. 1. The design of the digital display case

The digital display case is an important advance compared to conventional display cases from the viewpoint of simply displaying and manipulating various cultural assets. However, we need to prepare a 3D model to create the digital display, which is generally very time consuming even with the use of a 3D scanner, and we also need special techniques and tools to create high-quality content. This makes it difficult to place new objects in the digital display case at the museum, when there is no special operator to produce digital content. The curators of the art museum have high expectations for the digital display case, because they want to display a large number of cultural assets. This is possible with a digital display case; however, each art museum owns tens of thousands of cultural assets, and it is very difficult to prepare 3D models of all of them.

3 Virtual Transportation System

To solve the problem described in Section 2, we propose our virtual transportation system. Figure 2 shows the design of this system. This system consists of a display unit and a capture unit.

The display unit is used to exhibit virtual cultural assets. The basic functions and shapes of the display unit are identical to the previous digital display case.

The capture unit views real cultural assets and captures images of them. The shape of the capture unit resembles a conventional display case and can exhibit real cultural assets in a manner similar to conventional display cases. The capture unit has a function that can automatically generate images of arbitrary views by using image-based rendering (IBR) techniques. IBR techniques are image processing techniques

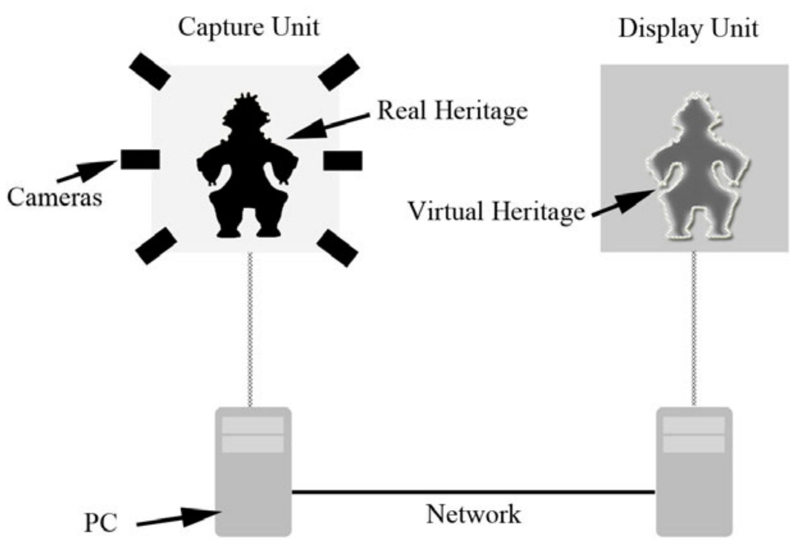


Fig. 2. The design of our virtual transportation system

that can generate the viewpoints of virtual cameras from the pictures of one or more real cameras. To capture input images for use of IBR techniques, multiple cameras surrounding the cultural assets are attached to the inside of the capture unit. Each unit is connected to a PC, and the two PCs are connected via network.

Following are the processes required to run this system:

1. The cultural assets are placed in the capture unit.
2. The capture unit captures multiple input images of the cultural assets.
3. The capture unit's PC sends the input images to the display unit's PC via network.
4. The display unit's PC computes the arbitrary view images from the received original images by using IBR techniques.
5. The images are displayed on the display unit.

The advantages of our system are it takes only a short time to reconstruct images of the assets, the processes are simple, and the output images are highly realistic. In terms of the reconstruction of the cultural assets, it takes only a few minutes to compute the arbitrary view images from the multiple input images taken by the cameras. Because process 1 entails only placing the cultural assets in the capture unit and processes 2–5 are computed automatically, the steps are simple to perform. The images of the virtual cultural assets in this system approximate the state of a real view better than that obtained by using a 3D scanner, because this system directly generates the images from the pictures of real cultural assets, in contrast with the view created by 3D scanners, which only approximately simulates the shape and lighting.

4 Experimental Result of the Prototype System

To indicate that the concept of our virtual transportation system is effective, we constructed a prototype system. By evaluating the prototype system, we verify that the capture unit can generate arbitrary view images in the same way as if 3D cultural assets were placed in the display unit.

In the prototype system, we selected View Morphing as the IBR technique [4]. View Morphing is an image morphing technique that can generate interpolated images of virtual cameras arranged on a line segment between the two cameras. One advantage of View Morphing is that it requires the use of fewer cameras than that required by other IBR techniques. Therefore, the number of cameras in the capture unit can be decreased, and the cameras can be arranged inconspicuously. However, because View Morphing can only generate the images of virtual cameras on the line segment, our experiment limits the transformation areas of the view on the display unit to horizontal directions. If we need to generate arbitrary viewpoints that include both horizontal and vertical directions, tri-view morphing [5] or trifocal transfer [6] may be required.

Figure 3 shows a replica of the cultural asset utilized in our experiment. The cultural asset, which is called DOGU, is a traditional ceramic figure in Japan. Figure 4 shows the setup for photographs of the DOGU. In these photographs, the DOGU was put on a turntable, and light was placed over the DOGU. The camera was set in a horizontal direction and aimed at the center of the DOGU. We captured the DOGU while the turntable was rotated by 30 degrees. View Morphing generated interpolation images that were each rotated by 3 degrees for the input images that were each rotated by 30 degrees. Figure 5 shows the pictures and the interpolation images.

Fig. 3. A replica of the cultural asset (DOGU)

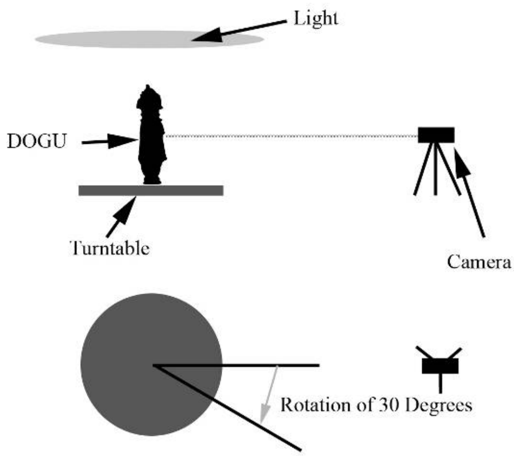


Fig. 4. The setup for photographs of the DOGU

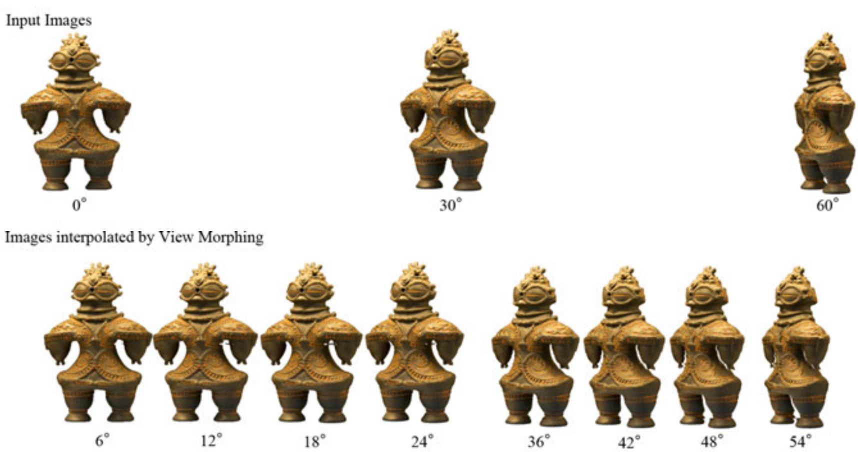


Fig. 5. Pictures of the DOGU and the images interpolated by View Morphing

In our experiment, the pictures of the DOGU were prepared as input images beforehand, and the interpolation images were pre-computed by View Morphing. The computation time was approximately 2 min to interpolate nine images from two input images at a resolution of 10 megapixels.

Figure 6 shows a demonstration of the prototype system. Figure 7 shows a mock-up of the capture unit, and Figure 8 shows the display unit. As a result of the demonstration, we obtained opinions that the virtual cultural asset was viewed similar to a 3D object, even though these images were interpolated by View Morphing by a two dimensional image processing technique. Therefore, we conclude that our proposed system can obtain views that are the same as those obtained using the previous digital display case, and that we can simply reconstruct virtual cultural assets by using IBR techniques.

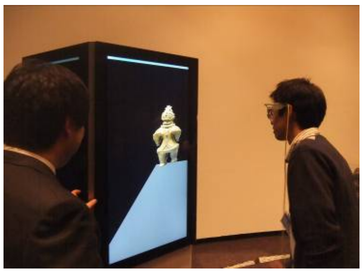


Fig. 6. The demonstration of our prototype system

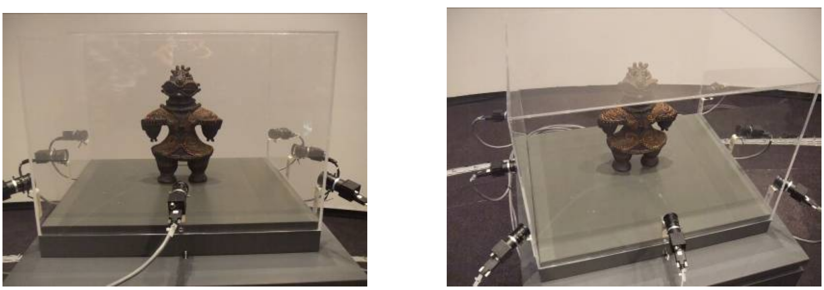


Fig. 7. A mock-up of the capture unit

Fig. 8. The virtual DOGU in the display unit

On the other hand, we encountered the following problem. The images interpolated by View Morphing included outlier which has a negative impact on the perceptual quality as shown in Figure 9. This is derived from the failure of scanline matching. Because visitors at art museums will pay close attention to the details of cultural assets, these outliers can be a serious problem. When we adopted IBR techniques other than View Morphing, broken points were also observed that were derived from the failure of scanline matching.

Fig. 9. An enlarged view of an interpolation image of the DOGU

5 Future Applications

This section introduces two future applications that our proposed system will achieve.

The first application is a remote exhibition system (Figure 10). This system aims to create an experience that reproduces the effect as if the visitors of an art museum were viewing the actual cultural assets in real time. This system can be achieved by displaying the cultural assets along with views of the actual surroundings in the display unit.

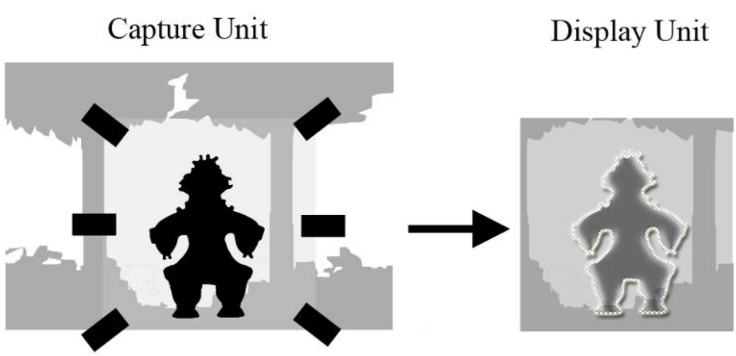


Fig. 10. A remote sightseeing system

The second application is a remote education system (Figure 11). This system predicates that anyone in the world can be educated by specialists without the physical transportation of cultural assets and specialists. This system can be achieved by techniques in which the specialists explain by manipulating real cultural assets in the capture unit, and where the display unit reconstructs the virtual cultural assets being manipulated.

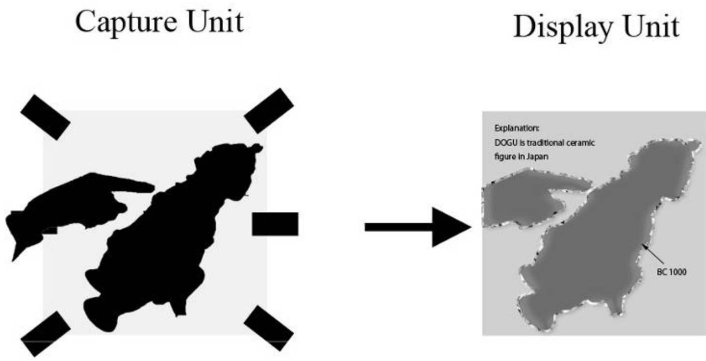


Fig. 11. A remote education system

6 Conclusion

This paper proposed a display case for art museums that utilizes computer graphics and IBR techniques. Our virtual transportation system consists of a display unit and a capture unit. The display unit can exhibit the virtual cultural assets as real, as if there are real cultural assets in the display case. The capture unit can automatically reconstruct the cultural assets in a short time by using IBR techniques. This system requires only the placement of the cultural assets into the capture unit. The experimental result of the prototype system shows that the images of the cultural asset interpolated by View Morphing were viewed as 3D objects.

As a future work, to solve the problem that the interpolation images created by IBR techniques cause outliers derived from the failure of scanline matching, we will further investigate the IBR technique used in this system. In addition, in order to achieve the remote sightseeing system and the remote education system, we will investigate techniques for acquiring surroundings and lighting and for performing real-time computations.

References

1. Ando, M.: Networked VR for Virtual Heritage. In: HCII 2003, vol. 3, pp. 1386–1390 (2003)
2. Ueno, R.: An Experience of Digital Cultural Assets at Museum. In: VSMM 2010 (2010)

3. Kajinami, T., Hayashi, O., Narumi, T., Tanikawa, T., Hirose, M.: Digital Display Case: The Exhibition Sysytem for Conveying the Background Information. In: SIGGRAPH 2010 (2010)
4. Seitz, S.M., Dyer, C.R.: View Morphing. In: SIGGRAPH 1996 (1996)
5. Xiao, J., Shah, M.: Tri-view morphing. Computer Vision and Image Understanding 96, 345–366 (2004)
6. Avidan, S., Shashua, A.: Novel view Synthesis in Tensor Space. In: CVPR 1997 (1997)

Part III
Virtual Humans and Avatars

Integrating Multi-agents in a 3D Serious Game Aimed at Cognitive Stimulation

Priscilla F. de Abreu[1], Luis Alfredo V. de Carvalho[2], Vera Maria B. Werneck[1], and Rosa Maria E. Moreira da Costa[1]

[1] Universidade do Estado do Rio de Janeiro
IME – Mestrado em Ciências Computacionais
Rua São Francisco Xavier 524- 6o andar – Bl. B,
CEP 20550-013 - Rio de Janeiro – RJ – Brasil
[2] Universidade Federal do Rio de Janeiro
COPPE - Programa de Eng. de Sistemas e Computação
Caixa Postal 68511 - CEP 21941-972
Rio de Janeiro - RJ - Brasil
priscillaf.uerj@gmail.com, LuisAlfredo@ufrj.br,
{vera,rcosta}@ime.uerj.br

Abstract. Therapies for cognitive stimulation must be developed when some of the cognitive functions are not working properly. In many applications there is a strong dependence on therapist's intervention to control the patient's navigation in the environment and to change the difficulty level of a task. In general, these interventions, cause distractions, reducing the level of user immersion in the activities. As an alternative, the inclusion of intelligent agents can help to alleviate this problem by reducing the need of therapist involvement. This paper presents a serious game that combines the technologies of Virtual Reality and Multi-Agent Systems designed to improve the cognitive functions in patients with neuropsychiatric disorders. The integration of different technologies and the modelling methodology are described and open new software development perspectives for 3D environments construction.

Keywords: Virtual reality, Multi-Agents Systems, Serious Games, Cognitive Stimulation.

1 Introduction

The cognitive functions are crucial for human life development. Therapies for cognitive stimulation must be developed when some of the cognitive functions are not working properly. These therapies can allow the recovery of basic functions such as attention and memory. Several strategies and technologies are being explored in this area and Virtual Reality (VR) applications is one that can be specially nominated. Virtual Reality includes advanced interface technologies, immersing the user in environments that can be actively interacted with and explored [10].

A wide range of everyday situations can be specifically developed through software designed for stimulation of the cognitive functions. Reasoning exercises can

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 217–226, 2011.

be repeated extensively and have their level of difficulty changed according to the patients' results.

Currently, these applications can be considered as "Serious Games" which are computer games with the goal of education and/or construction of concepts [9]. These games allow the simulation of real-world situations, providing training activities that stimulate cognitive functions and psychomotor skills.

Recently, several experiments in this area have had positive results, which are stimulating new research [8]. However, in many applications there is a strong dependence on therapist's intervention to control the patient's navigation, to change the level of difficulty of a task, to go back to an earlier stage, or to control the order of activities in the synthetic environment. These interventions, however needed, cause distractions, reducing the level of user immersion in the activities. As an alternative, the inclusion of intelligent agents can help to alleviate this problem by reducing the need of therapist involvement.

This paper aims at presenting a serious game that combines the technologies of Virtual Reality and Multi-Agent Systems designed to improve the cognitive functions in patients with neuropsychiatric disorders. The agents control user's actions register in the system and propose tasks according to his level of performance.

The game integrates a group of agents which will be responsible for managing and monitoring the performance of individuals in the tasks. The game's levels of difficulty will be controlled by agents. Agents respond to user action and build a personal trajectory of activities. The game is composed by two different rooms: one aimed to stimulate attention and concentration and the other related to memory. These rooms contain furniture and decorations similar to those used in family homes. The first game shows a room with three shelves, one beside the other, with different objects on the shelves. Some objects similar to those in the bookshelf will be randomly shown on a frame and the patient must walk to one of the shelves and click on the corresponding object. The second one presents an object associated with a number and the user must click on as many objects as were requested.

People with different disabilities can use this game, especially those who must receive stimuli of spatial attention, such as people with unilateral spatial neglect.

The game development process explored specific methodologies for agent modelling and implementation.

This work is organized into 5 sections. Section 2 gives an overview of the main themes related to this work, and Section 3 describes the adopted methodology for the game development. Section 4 presents the game prototype. Section 5 concludes the work and presents future research directions.

2 General Concepts

After an injury the brain is able to rearrange its connections and consequently, the functional recovery [7]. Based on the brain plasticity ability, the cognitive rehabilitation emerges as a strong ally for cognitive function recovering of individuals who suffered some brain damage. In general, the exercises produce a specific reorganization in various levels of neural connections, increasing the recovery process of cognitive and motor functions.

The cognitive rehabilitation functions such as visual perception, attention and memory, the motor skills explore the 3D virtual environments potential supported by different theoretical and practical approaches [8]. In general, these applications are developed in game format to motivate the patient to do the tasks. Another possibility that has been considered in this area is the use of commercial games that are used for specific purposes. In this case, the games are known as “Serious Games”. The "Serious Games" offers activities that help users in absorbing concepts and psychomotor skills. Generally this term is used to highlight games that aim to provide, besides entertainment, experiences related to education and training [3], [9].

The Virtual Reality technology has been widely used in these games, providing opportunities to offer some situations closer to the real world. Burdea and Coiffet [10] defines the applications of Virtual Reality as a three-dimensional virtual environments that presents real-time graphics rendered by a computer, in which the user, via body position sensors or user-input devices, controls the viewpoint or the orientation of displayed objects.

In the area of cognitive rehabilitation through 3D environments, several research groups have developed and tested products for different level of disabilities. Gamito et al. [11] created a virtual environment consisting of a small town with digital robots, several buildings and a mini market where the user can move freely. The tasks stimulate the executive functions associated with activities of daily living such as personal care and identification of routes, among others. Meijer et al. [4] present a virtual environment composed of a supermarket with several groceries sections with the objective to verify the user's spatial learning by assessing the degree of accuracy that he learns the route and layout of the environment. Attree et al. [5] describe an environment where the user should observe the objects arranged around the room and try to find a specific toy car among them, stimulating memory and attention.

Despite the many research efforts addressing the development of these environments, it does not have control over the users' navigation, nor propose the tasks in an automated way. In many applications there is a strong dependence on the therapist's intervention to control the patient’s navigation, to change the level of difficulty of a task, to go back to an earlier stage, or to control the order of activities in the synthetic environment. These therapists’ interventions can reduce the level of user immersion in the simulation. As an alternative, the inclusion of intelligent agents could help to alleviate this problem, by reducing the need of therapists’ involvement.

2.1 Agents

Multi-agents systems applications have been increasing in different areas. An agent can be considered an autonomous system seeking different ways to reach pre-established goals in a real or virtual environment [12]. Another definition describes an agent as a system capable of perceiving the information from their environment through sensors and acting through actuators [13]. Every agent must have autonomy, which means that an agent has the ability to manage its internal state and its actions to achieve their goals without human intervention. An agent may have other characteristics, which have a degree of importance depending on the area in question, reactivity, adaptability, communication, mobility, etc.

Agents are classified regarding intelligence [12] into Reactive Agents, Deliberative and Cognitive Agents or Hybrid Agents.

Reactive Agents are simple agents based on simple event-response model, reacting to environmental changes. These agents have no memory and thus are unable to plan future actions [14]. The idea of this architecture is that a global intelligent behavior is achieved by the interaction of several simple behaviors.

Deliberative and Cognitive Agents are based on models of human organization such as communities and hierarchies. These agents can interact with other agents using complex languages and messaging protocols. They have explicit representation of the environment, community members and can reason about actions taken in the past and plan future actions. In general, have high computational complexity. Among the architectures developed for the creation of these agents, there is the architecture Belief, Desire and Intention (BDI) which is based on mental states: beliefs, desires and intentions [13].

Hybrid Agents have reactive and cognitive architectures components. They are not purely reactive or cognitive [13].

A multi-agent system is composed of two or more agents who have a set of skills and plans in order to achieve their goals [15].

2.2 Agents in 3D Environments

In the area of Virtual Reality (VR) the use of MAS is still new however we can find some applications in areas of education and training. Some of these applications are described below.

STEVE (Soar Training Expert for Virtual Environment) is an animated pedagogical agent embedded in a 3D simulation system designed to assist students in naval training [16]. STEVE is considered an intelligent tutoring system that integrates methods from three research areas: computer graphics, intelligent tutoring systems and agent architecture [17].

Active Worlds is a business and personal applications platform for online distribution of interactive 3D content in real time. Personal applications can build personal learning environments, as the River City project, a virtual environment for science classes in high school. In this virtual world, students travel in time bringing their skills and technologies of the XXI century to solve problems of the nineteenth century. The students work with research teams to help a city to understand what the causes of the residents' sickness are. The students, who are represented by avatars, interact through a chat with the locals, which are software agents that respond to questions from students [18].

3 Game Development Methodology

The game development methodology was organized into four phases: (i) study of rehabilitation process using Virtual Reality, (ii) study of the technology of virtual environments and multi-agent systems, (iii) modelling and developing the game, (iv) evaluation the game for the rehabilitation process.

In the first phase we studied the levels of cognitive development of an individual, the consequences of brain injuries, the cognitive rehabilitation process and the experiences with the use of VR as a tool to support the rehabilitation process. A physician gave the initial software requirements, discussing the tasks and the difficulty levels.

At the second phase we studied and analyzed the technologies adopted for development of virtual environments and multi-agent systems. For this, we conducted a survey of these tools prioritizing those that were free. From that, the tools were chosen and tested. A development of a prototype involving the integration of these tools was also performed. We did not find in the literature works that describe how to make this integration.

Then in the third phase the system was modelled and the first prototype was implemented. After that we had a meeting with a physician to assess the tasks, the level of difficulties and the game interface.

In the last phase some experiments will be made with a small group of patients to evaluate the game usability. Then a new evaluation will be conducted with experts.

3.1 Game Modelling

For modelling the game we conducted an interview with an expert in the application area to get the initial software requirements. Then a lexical catalogue using LAL (Language Extended Lexicon) [26] was specified to register and document the requirements [25].

The goal oriented framework i* (i-star) [24] was adopted for modelling the game and it models the contexts based on the dependency relationships among actors. Actors depend on each other for goals to be achieved, for resources to be provided, for tasks to be performed, and for softgoals (non-functional requirements) to be satisfied. This framework defines two kinds of models: the Strategic Dependency (SD) model and the Strategic Rationale (SR) model.

The Strategic Dependency SD model consists of a set of nodes and links defining dependency between two actors. One actor (called depender) depends on one or more actors (called dependees) through a dependency (either an end node or a component node called dependum). A dependum may be one of four types: a softgoal, a task, a resource or a goal [24].

The Strategic Rationale (SR) model expresses an actor's reasoning. Rationales are modelled through means-ends relationships, task decompositions, and softgoal contributions. The means-ends relationship represent relationships like a goal to be achieved, a task to be done, a resource to be produced, or a softgoal to be satisfied named as the "end" and the "means" alternative to achieve them. The tasks decomposition are associated with 'how to do something' and can be modelled in terms of its subcomponents decomposition which can be: goals, tasks, resources and/or soft goals [24].

We defined four agents that are responsible for planning patient care, controlling the interactions in games, analyzing the performance of the patient and setting up the environment. The SD diagram in Figure 1 helped to identify these agents and users and their relationships. For example the planner depends on the analyser to change the protocol and the analyzer on the other hand depends on the controller to provide the patient game interaction history.

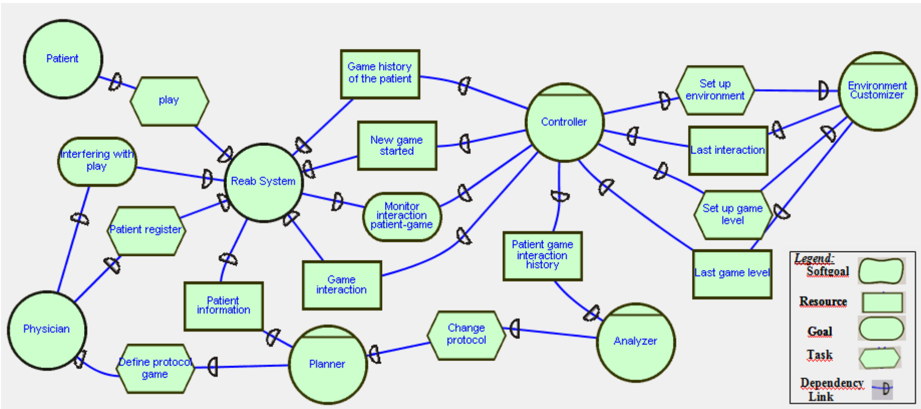


Fig. 1. Strategic Dependency (SD) model

The i* framework in the SR Diagram details the rational of each agent by describing how the goals are archived, the task performed, the softgoal satisfied and the resources obtained. The Figure 2 is the SR game for the agent Analyser which is responsible to archive the goal "performance evaluation of the patient". This goal is satisfied by the task "assess patient performance" that is decomposed into three tasks (capturing treatment protocol, capturing game history and comparing the expected performance) and the goals "rated performance". The task "capturing game history" depends on the Control Agent to provide the information "patient game interaction history". The goal "rated performance" is satisfied by performing one of the tasks "continue in the same plane" or "suggest change of plan". This last task will provide the resource "change protocol".

4 The Game Prototype

A first prototype was implemented to validate the requirements and verify the game usability with an expert in the CR area. It was built based on i* models defined before.

The game environment integrates the technologies of Virtual Reality and Multi-agent Systems and aims to stimulate the cognitive functions such as attention, concentration and memory. The levels of difficulty of the proposed activities are controlled by agents.

For the implementation of the agents and 3D environment we studied several languages and frameworks, verifying the technological compatibility among them. Thus, various combinations were tested to achieve this goal [19]. Among the technologies studied, we adopted the following languages: X3D (eXtensible 3D) [20], JAVA [21] and the framework JADE (Java Agent Development Framework) [22]. To integrate these three applications we used the NetBeans Platform [23]. The agents' behavior is based on the Cognitive Model [12]. The agents have goals and desires that are combined to control the difficulty level of the game.

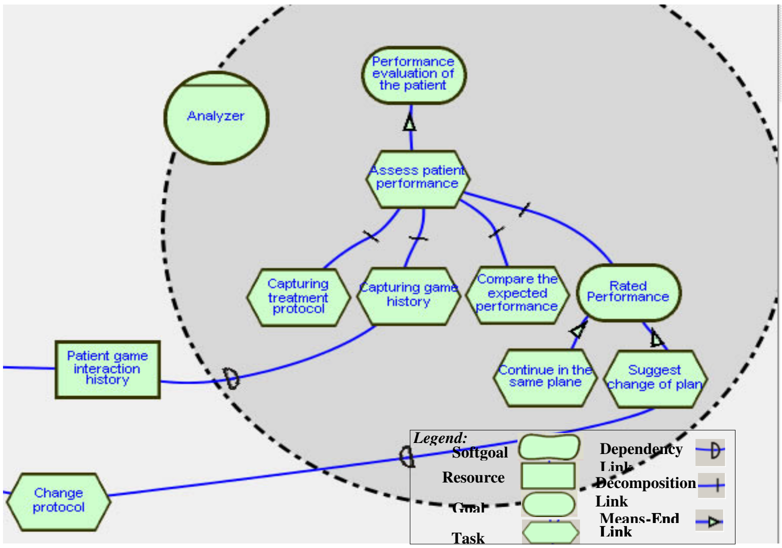


Fig. 2. SR Model of Analyser Agent

The games are distributed into two rooms. The first game has bookshelves, with different objects on them. Randomly objects are shown on a table and the patient must walk to one shelve and click on the similar one. An agent will monitor the time and the user interactions with the environment, controlling the rights and wrongs answers, changing the difficulty level of the tasks according to their performance.

Another game will occur in another room (a kitchen) and will comprise several objects that are part of day by day life of an individual. In this activity, the degree of difficulty is greater than the initial exercise described above, because the patient will deal with two variables: an object and a number 'n', which will be presented at the scene. The user must click on 'n' similar objects. The game's difficulty level will rise by increasing the number of objects to be selected and by the introduction of distracters in the scene as a sound or a blinking object. The second game scene is in the Figure 3.

Aiming at analyzing some aspects of the prototype we made an evaluation with a psychiatrist, based on a script that follows the concepts of consolidated methodologies. The psychiatrist evaluated this first prototype by analyzing those aspects: Navigation facility, Learning facility, Response time, Realism of scenes, Pleasantness of the scenes, Adequacy of the objects in the tasks, Matching Colours and Adequacy of the difficulty level of the game.

Fig. 3. The kitchen where the second level game will be developed

5 Conclusions

Democratizing access to new treatment practices for neuropsychological disorders has been the target of several research groups. In Brazil, we observe a growing interest in this area. In this sense, we need new software and new models of treatment, where patients may have more unrestricted access to the exercises and the therapist can monitor the result at distance, reducing the need for constant face-to-face contacts. Some types of exercises could be done at home, expanding the possibilities of rehabilitation. Thus, in this case, the virtual environment must have some mechanisms to control user navigation and generate automatic reports to the therapist. The Virtual Reality technology is widely used to give cognitive and motor stimulus for patients with different impairments.

However, the development of such software depends on knowledge of different fields of expertise. The modelling, the design, the implementation, the programming of intelligence and the assessment are examples of tasks present in the development process.

This paper presented the initial results of a project that has two objectives. The first one is associated with the technical questions related to the intelligent agents modelling that will be in the 3-D virtual environment. The other aims at finding the combination of technologies to support the integration of intelligent agents within the 3-D environments.

To develop this environment we adopted a methodology composed by 4 steps supported by a multidisciplinary team. We considered that we found a good solution for this problem. The integration of X3D, JAVA, JADE and NetBeans created a new combination of technologies that worked efficiently in our initial tests.

Acknowledgments. This study is supported in part by the Instituto Nacional de Ciência e Tecnologia - Medicina Assistida por Computação Cientifica (Medicine assisted by Scientific Computation), Brazil.

References

1. Census. Brazilian demographic Census (2000), `http://www.ibge.gov.br/home/presidencia/noticias/27062003censo.shtm` (visited on January 2011)
2. Berger-Vachon, C.: Virtual reality and disability. Technology and Disability 18, 163–165 (2006)
3. Machado, L. S., Moraes, R.M.,. Nunes, F. L. S, Costa, R. M. E. M.: Serious Games based on Virtual Reality for Medical Education, Serious Games Baseados em Realidade Virtual para Educação Médica. Revista Brasileira de Educação Médica (in press)
4. Meijer, F., Geudeke, B.L., Broek, E.L.V.D.: Navigating through Virtual Environments: Visual Realism Improves Spatial Cognition. Cyberpsychology & Behavior 12(5) (2009)
5. Attree, E.A., Turner, M.J., Cowell, N.: A Virtual Reality Test Identifies the Visuospatial Strengths of Adolescents with Dyslexia. CyberPsychology & Behavior 12(2) (2009)
6. Josman, N., Klinger, E., Kizony, R.: Performance within the virtual action planning supermarket (VAP-S): an executive function profile of three different populations suffering from deficits in the central nervous system. In: VII International Conference Disability, Virtual Reality & Associated Technologies (ICDVRAT), Portugal, Maia (2008)
7. Sohlberg, M.M., Mateer, C.: Cognitive Rehabilitation: An Integrated Neuropsychological Approach. Guilford Publication, New York (2001)
8. Rizzo, A.A., Bowerly, T., Buckwalter, J.G., Klimchuk, D., Mitura, R., Parsons, T.D.: A Virtual Reality Scenario for All Seasons: The Virtual Classroom. CNS Spectr. 11(1), 35–44 (2006)
9. Larsson, P.A., Broeren, J., Bellner, A.-L., Fogelberg, M., Goransson, O., Goude, D., Johansson, B., Pettersson, K., Rydmark, M.: Exploration of computer games in rehabilitation for brain damage. In: Proceedings VII International Conference Disability, Virtual Reality & Associated Technologies (ICDVRAT), Maia, Portugal (2008)
10. Burdea, G., Coiffet, P.: Virtual Reality Technology, 2nd edn. John Wiley & Sons, New Jersey (2003)
11. Gamito, P., Oliveira, J., Pacheco, J., Morais, D., Saraiva, T., Lacerda, T., Baptista, A., Santos, N., Soares, F., Gamito, L., Rosa, P.: Traumatic brain injury memory training: a virtual reality online solution. In: Proceedings VIII International Conference Disability, Virtual Reality & Associated Technologies, Chile (2010)
12. Wooldridge, M.J.: An Introduction to Multi-Agent Systems. John Wiley and Sons Limited, Chichester (2009)
13. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach, 3rd edn. Prentice Hall, Englewood Cliffs (2009)
14. Bigus, J.P., Bigus, J.: Constructing Intelligent Agents using Java, 2nd edn. Professional Developer's Guide Series. Wiley, Chichester (2001)
15. Gago, I.S.B., Werneck, V.M.B., Costa, R.M.: Modeling an educational multi-agent system in maSE. In: Liu, J., Wu, J., Yao, Y., Nishida, T. (eds.) AMT 2009. LNCS, vol. 5820, pp. 335–346. Springer, Heidelberg (2009)
16. Rickel, J., Johnson, W.L.: STEVE: A Pedagogical Agent for Virtual Reality. In: International Conference on Autonomous Agents (1998)
17. Ramírez, J., de Antonio, A.: Automated planning and replanning in an intelligent virtual environments for training. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) KES 2007, Part I. LNCS (LNAI), vol. 4692, pp. 765–772. Springer, Heidelberg (2007)

18. Ketelhut, D.J., Dede, C., Clarke, J., Nelson, B., Bowman, C.: Studying situated learning in a multi-user virtual environment. In: Baker, E., Dickieson, J., Wulfeck, W., O'Neil, H. (eds.) Assessment of Problem Solving using Simulations. Lawrence Erlbaum Associates, Mahwah (2007)
19. Costa, R.M.E.M., Mendonca, I., Souza, D.S.: Exploring the intelligent agents for controlling user navigation in 3D games for cognitive stimulation. In: 8th International Conference on Disability, Virtual Reality and Associated Technologies, vol. 1, pp. 1–6 (2010)
20. Brutzman, D., Daly, L.: X3D: 3D Graphics for Web Authors. Morgan Kaufmann Publishers, San Francisco (2007)
21. JAVA, http://www.java.com/pt_BR/
22. JADE, http://jade.tilab.com/
23. Netbeans, http://netbeans.org/
24. Yu, E., Liu, L.: Modelling trust for system design using the *i** strategic actors framework. In: Falcone, R., Singh, M., Tan, Y.-H. (eds.) AA-WS 2000. LNCS (LNAI), vol. 2246, pp. 175–194. Springer, Heidelberg (2001)
25. Cenários e Léxicos (C&L), http://pes.inf.puc-rio.br/cel/
26. Leite, J.C.S.P., Franco, A.P.M.: A Strategy for Conceptual Model Acquisition. In: Proceedings of 1th IEEE International Symposium on Requirements Engineering, pp. 243–246. IEEE Computer Society Press, San Diego (1993)

Automatic 3-D Facial Fitting Technique for a Second Life Avatar

Hiroshi Dohi[1] and Mitsuru Ishizuka[2]

[1] Dept. Information and Communication Engineering, Graduate School of Information Science and Technology, University of Tokyo
[2] Dept. Creative Informatics, Graduate School of Information Science and Technology, University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
dohi@mi.ci.i.u-tokyo.ac.jp

Abstract. This paper describes an automatic 3-D facial fitting technique for a Second Life avatar. It is often difficult to create an original avatar by yourself that resembles a real person. In Second Life, the combinations of system-defined parameters deform the shape of the avatar, and we can't control each vertex directly. It needs to encode the deformation information into many parameters. As a reference target, we make use of MRI data that scans the real face. It is stable and not affected by lighting and diffused reflection. In our experiments, we picked up 428 vertices on the base model for facial fitting. Using the iteration technique, more than 50% of vertices are just on the reference targets, and more than 85 % are within +/- 3mm of errors.

1 Introduction

A 3-D virtual environment is expected as one of post-WWW media interface. Many users share the 3-D world, and interact with each other through their avatars.

The avatar is a digital persona in a 3-D virtual world. There are two types of avatars. One is the avatar that has alternate identity and another is the avatar that resembles a real character. The former is the "anonymous" avatar. It can hide real name and protect user's privacy in real life. And the latter, "identity" may have special meaning to others. The avatar is associated with the person in real life.

It has been discussed for a long time about the avatar that resembles a famous character and "uncanny valley" hypothesis. The realistic face, of course, is not necessarily critical for avatar communication. However, in real life the famous character is often used for advertising, e.g. a movie star, a celebrity, an athlete, a politician, and so on.

The human has superior discrimination capability for human faces. It is not difficult for you to find your friend out of the crowd. Even the avatar that resembles your closest friend may have meaning in your local community group in the 3-D virtual world.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 227–236, 2011.

It is often difficult to create an original avatar by yourself that resembles a real person.

For example, Linden Lab's Second Life [9] has gathered up more than 18 million registered users (called "residents") all over the world. One of salient features on Second Life is high-quality computer graphics. We can purchase various types of realistic objects, includes apparels, avatar accessories, and even avatar appearance, in Linden$ currency at the marketplace in the Second Life world [10]. It is easy to become a different avatar. However it is difficult to create the avatar that resembles you by yourself, although Second Life offers an avatar appearance tool to all residents. The tool is simple to use, but it requires skills of both art and computer graphics for creating the fascinating avatar.

In this paper, we have empirical study of automatic 3-D facial fitting technique for a Second Life avatar. It needs to encode deformation information to many visual parameters.

2 Avatar Appearance in Second Life

In Second Life, many avatars live in the 3-D virtual environment. Each user can watch the world from arbitrary viewpoint.

Table 1. Shape parameters of the appearance tool in Second Life

Sub categories		Parameters
Body*	3	Height, Body Thickness, Body Fat
Head	11	Head Size*, ***Head Stretch, Head Shape, Egg Head, Head Length,*** Face Shear*, ***Forehead Angle, Brow Size, Upper Cheeks, Lower Cheeks, Cheek Bones***
Eyes	11	***Eye Size, Eye Opening, Eye Spacing, Outer Eye Corner, Inner Eye Corner, Eye Depth, Upper Eyelid Fold, Eye Bags, Puffy Eyelids,*** Eyelash Length*, Eye Pop*
Ears*	4	Ear Size, Ear Angle, Attached Earlobes, Ear Tips
Nose	11	***Nose Size, Nose Width, Nostril Width, Nostril Division, Nose Thickness, Upper Bridge, Lower Bridge, Bridge Width, Nose Tip Angle, Nose Tip Shape,*** Crooked Nose*
Mouth	9	***Lip Width, Lip Fullness, Lip Thickness, Lip Ratio, Mouth Position, Mouth Corner, Lip Cleft Depth, Lip Cleft,*** Shift Mouth*
Chin	9	***Chin Angle, Jaw Shape, Chin Depth, Jaw Angle, Jaw Jut, Jowls, Chin Cleft, Upper Chin Cleft, Chin-Neck***
Torso*	12	Torso Muscles, Neck Thickness, Neck Length, Shoulders, Breast Size, Breast Buoyancy, Breast Cleavage, Arm Length, Hand Size, Torso Length, Love Handles, Belly Size
Legs*	8	Leg Muscles, Leg Length, Hip Width, Hip Length, Butt Size, Saddle Bags, Knee Angle, Foot Size

* Not evaluated in our experiments for facial fittings.

The Second Life system uses a server-client model. The client (viewer) has the base 3-D wire frame model of the avatar. The position of each vertex is not transferred. Instead, the server distributes packets for visual parameters to all clients. It defines 218 visual parameters.

The Second Life official viewer includes some built-in tools. An avatar appearance tool is designed for modifying face / body shapes and creating clothes.

In the avatar appearance tool, “Body Parts” consists of 4 categories, i.e., ***Shape***, ***Skin***, ***Hair***, and ***Eyes***. The ***Shape*** category has 9 sub categories.

Table 1 shows shape parameters of the appearance tool.

The appearance tool can control various parameters listed on Table 1 by manipulating some slide bars. It is easy to create your original avatar.

However, it is more difficult to create the avatar that resembles a real person than you thought. Various techniques have been proposed for 3-D facial fitting. Blanz showed 3-D face reconstructions using a 3-D morphable model and optic flow algorithm [1,2,8].

In Second Life, there are many constraints for avatar deformation.

- We can’t control any vertices of the model directly.
- A single slide bar controls one or more visual parameters. It moves a group of vertices simultaneously. Each vertex moves in the straight line along with the slide bar value. Many vertices in the group move to different directions with different speeds in 3-D space.
- It is system-defined that which vertices are included in the group. The moving direction and speed of each vertex are also system-defined.
- The groups of vertices are overlapped. More than 40 slide bars may move a same vertex.
- There are about 50 slide bars for facial shape control. The range is usually “0 to 100” (or “–50 to +50”) and it is restricted to be integer values. As a result, it can create more than 10^{100} ($=101^{50}$) facial shapes theoretically. It may include duplicated shapes.

The new position of the vertex $\boldsymbol{V_n}$ is determined by following equations.

$$\vec{V}_n = f(\vec{v}_n, p_1, \ldots, p_m)$$
$$= \vec{v}_n + \sum_{i=1}^{m} w_i(p_i) \cdot \vec{d}_{i,n}$$

$\boldsymbol{v_n}$ is the base position of the vertex n. $\boldsymbol{p}$ is a morphing parameter and it is restricted to be integer values. $\boldsymbol{m}$ is the index of morphing parameters for facial control. $\boldsymbol{d}$ is a morphing vector. Each vertex has own vector on each morphing parameter. $\boldsymbol{w_i(p_i)}$ returns the weight (coefficient) of the morphing vector. Each morphing parameter has own range of $\boldsymbol{w}$. Both the range of $\boldsymbol{w}$ and the morphing vector are system-defined.

3 Facial Fitting Technique

3.1 MRI Data

As the reference target for facial fitting, we make use of the MRI data that scans the real face. "MRI (Magnetic Resonance Imaging)" is a medical imaging technique to visualize detailed internal tissues of the human body.

MRI data has following features.

- It is non-invasive measure, and it is safe for human body since it doesn't use ionizing radiation unlike X-rays technique.
- It can discriminate internal tissues. It can capture even the shape of the rear head covered with hair.
- It has no blind spots and no occlusion problem.
- MRI data is stable since it isn't affected by diffuse reflection.

A MRI system is very expensive and requires large space because it makes use of strong magnetic field. There are, however, unexpectedly many MRI units for medical use. According to the "OECD Health Data 2010," [7] there are 43.1 MRI units (and 97.3 computed tomography scanners) par million populations in Japan in 2008. And the number of MRI units increases twice from 2000 to 2008 across OECD (Organisation for Economic Co-operation and Development) membership countries.

The problem is that it is difficult to get raw image data from a medical institution. MRI data is one of medical data that includes private information, and then it is strictly controlled to bring it out.

3.2 Basic Algorithm

As the reference facial model, we used numerical human-body models developed by National Institute of Information and Communications Technology (NiCT), JAPAN [6]. These models are developed for the research on high-precision dosimetric

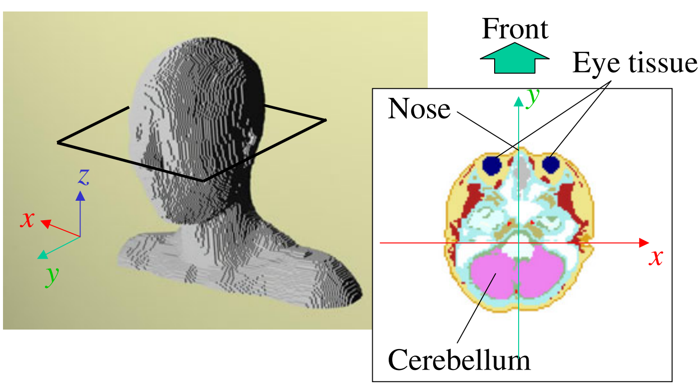


Fig. 1. A sliced MRI image

technology. These are voxel (volumetric pixel) data based on accumulated MRI images. With post processing, internal tissues are already labeled. The spatial resolution is 2 x 2 x 2 mm. The voxel size is 320 x 160 x 804 for the female. It is inhibited to specify the test subjects under the license agreement.

Fig.1 shows an example of the MRI image sliced on the ***x-y*** plane.

We make fitting the base model of the avatar to the reference target by combinations of system-defined parameters.

Our problem is to determine a set of morphing parameter values ($p_0, p_1, \ldots p_m$) that achieve the lowest errors below. ***p*** is integer value (from 0 to 100), and the integer programming problem is much more difficult.

$$Errors = \sum_{i=1}^{n} | Vi_y - D(Vi_x, Vi_z) |$$

D is depth information of the reference face. The set of parameter values ($p_0, p_1, \ldots p_m$) moves v_i to the new position V_i ***(x, y, z)***, and ***D*** depends on the position ***(x, z)*** of ***Vi***.

In order to decide parameters ($p_0, p_1, \ldots p_m$), we adopted a simple hill-climbing algorithm.

1. Choose one of morphing parameters ***p*** in random order.
2. Change the parameter value from 0 to 100, and calculate each error value.
3. If error value is lower, update the parameter value.
4. Iterate above steps until all parameters are not changed.

3.3 Optimization

The basic algorithm above needs to check parameters over and over again, since other parameters may move the same vertex.

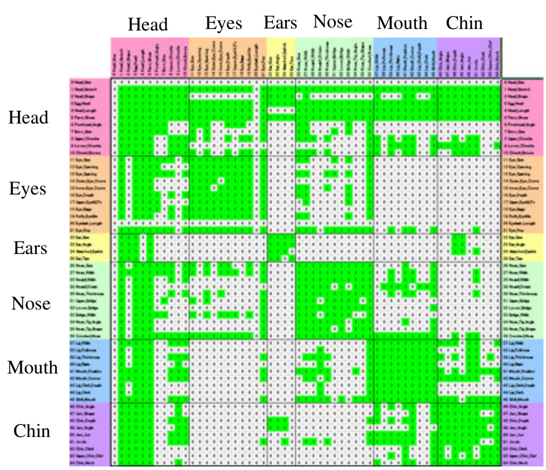


Fig. 2. Dependency matrix for morphing parameters

In order to decrease redundancy of the basic algorithm, we analyzed dependency relation of morphing parameters from avatar definition files, and made a dependency matrix

Fig. 2 shows the dependency matrix for morphing parameters.

The row of the matrix is the parameter that moves some vertices, and the column is the affected parameters. It is the symmetric matrix.

We use a bit vector in order to manage parameters efficiently. "The bit is 1" means that when we control the corresponding parameter it may make error value lower. The initial state of the bit vector is all 1. We choose one of the parameter that the bit is 1. Assume the current bit vector is "00110011." ($\boldsymbol{b_2} = \boldsymbol{b_3} = \boldsymbol{b_6} = \boldsymbol{b_7} = 1$) It means we have 4 candidates, i.e., parameters $\boldsymbol{P_2}$, $\boldsymbol{P_3}$, $\boldsymbol{P_6}$, and $\boldsymbol{P_7}$. For example, we choose $\boldsymbol{P_3}$, and clear the bit ($\boldsymbol{b_3} = 0$). Then we change the $\boldsymbol{P_3}$ value from 0 to 100, and calculate each error value. If error value is larger or same, the $\boldsymbol{P_3}$ value is unchanged and the bit vector shows next candidates. If lower, the $\boldsymbol{P_3}$ value is updated. Then it adds the dependency bit vector (the column line for $\boldsymbol{P_3}$ of the dependency matrix.) The column line "01101000" for $\boldsymbol{P_3}$ means that $\boldsymbol{P_3}$ affects $\boldsymbol{P_1}$, $\boldsymbol{P_2}$, and $\boldsymbol{P_4}$. When the bit vector is all 0 after iterations, we get the set of parameters that error value is one of local minima.

In order to escape local minima, it replaces some parameter values with random ones, and adds the dependency bit vector for the parameter to the current bit vector. Then it iterates above steps until the current bit vector becomes all 0.

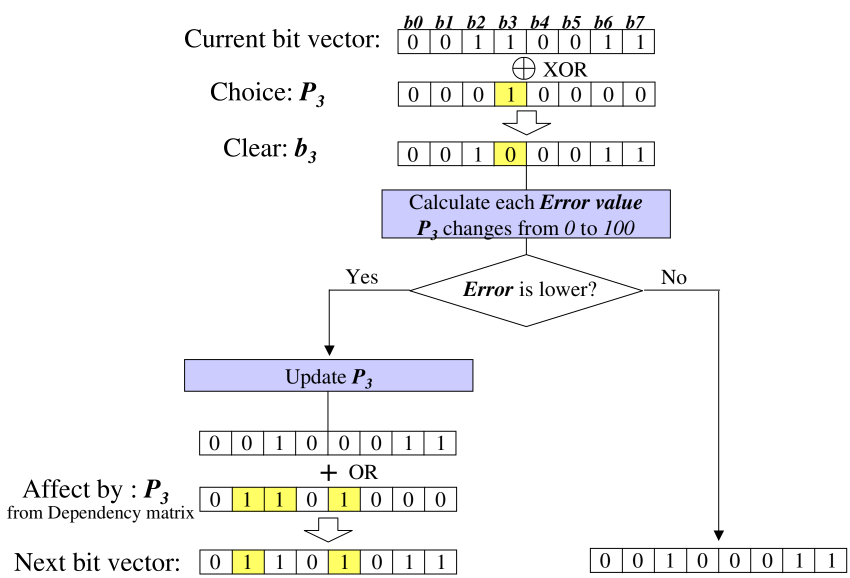


Fig. 3. Optimization of the basic hill-climbing algorithm

4 Results and Discussion

4.1 Facial Modeling Using MRI Data

It is useful to make use of MRI image. MRI has strong relation with various visualization techniques on computer graphics field, e.g. 3-D image reconstruction, voxel rendering, etc. However its data in itself has received little attention so far for facial modeling and animation, although MRI can scan a human body in detail.

Fig. 4 shows an example of the 3-D image reconstructed from MRI data.

The shape of the face is extracted well, even the rear head. The remarkable feature is that MRI data is stable. It is easy to extract outlines from the MRI voxel data.

It may seem that the shape is a coarse grid at a glance, but it is negligible since the size of the unit cube is small, 2 x 2 x 2 (mm) in our experiments.

3-D digitizers with laser (or optical) scan have been used for facial modeling. The 3-D digitizer generates a surface model, and the resolution is generally more accurate (<1 mm) than that of MRI. However, it can't capture hair data because strands often diffuse and reflect sensor beam. Other data, e.g., eyeball, eyebrow, beard and mustache, etc. are sometimes missing or wrong. It may require post processing for data modification manually.

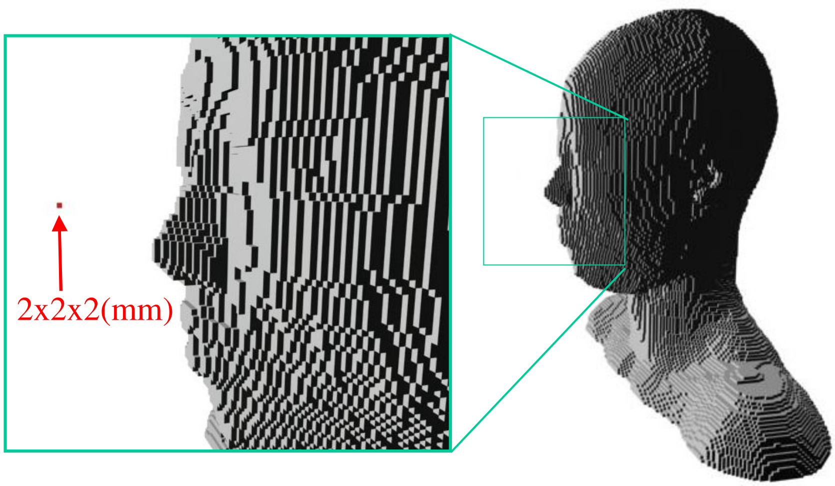


Fig. 4. 3-D image reconstructed from MRI data

4.2 3-D Facial Fitting for Second Life Avatar

Occluded (or overlapped) vertices (covered by other surface) on the base facial model may cause improper results, because our experiments use depth information. We removed excess vertices that could cause problems, e.g. contours of the side face, around of eyes, ears, nostril, mouth / lip, and rear head. The head part of the base model is composed of 1132 vertices (without LOD vertices), and we picked up 428 vertices for our evaluation.

In our experiments, we don't make fitting the rear head because Second Life has only a few parameters for deforming it. The rear head is often covered with hair. Hair with various hairstyles is another item.

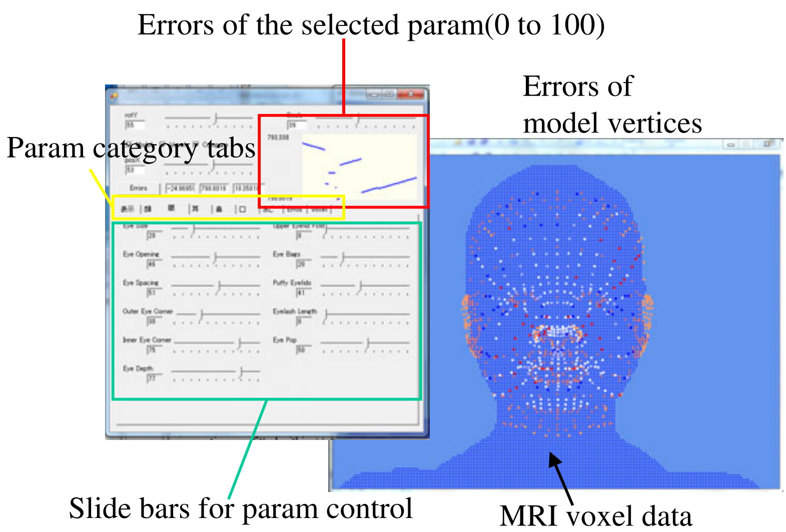


Fig. 5. A screen shot of our evaluation system

Fig. 5 shows a screen shot of our evaluation system. There are two windows. One is the parameter controller window (left), and another is the error display window (right).

The parameter controller has parameter category tabs and many slide bars. These have same functions as the avatar appearance tool on the Second Life viewer.

In our system, when we click on the label of the slide bar, it calculates all error values that the selected parameter changes from 0 to 100 automatically. Then, it displays those results as a graph at the upper right of the window, and the slide bar is set to the parameter value that gives the lowest error value.

It also has the function that iterates until all parameters become unchanged. Finally, we can get the set of parameter values.

The error display shows the position and depth error of each vertex.

The shape of head (small blue dots) is MRI voxel data. White boxes show model vertices that are just fitting. Red boxes have plus errors, and blue boxes have minus errors. Yellow boxes are model vertices that not evaluated in our experiments.

Fig. 6 is the side view. It finds fitting is well (white boxes) along the outline from forehead to chin. On the other hand, the lower part of the rear head is apart from the reference model since there is no morphing parameter. These vertices are not evaluated (yellow boxes).

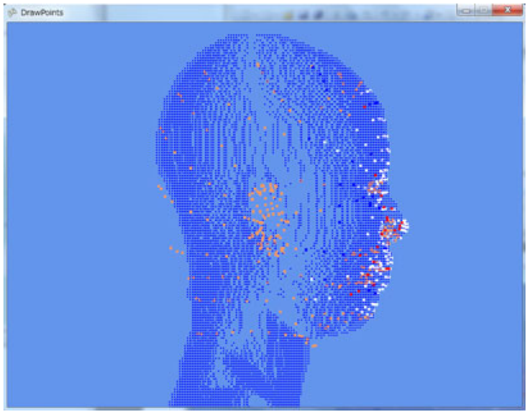


Fig. 6. Fitting results (side view)

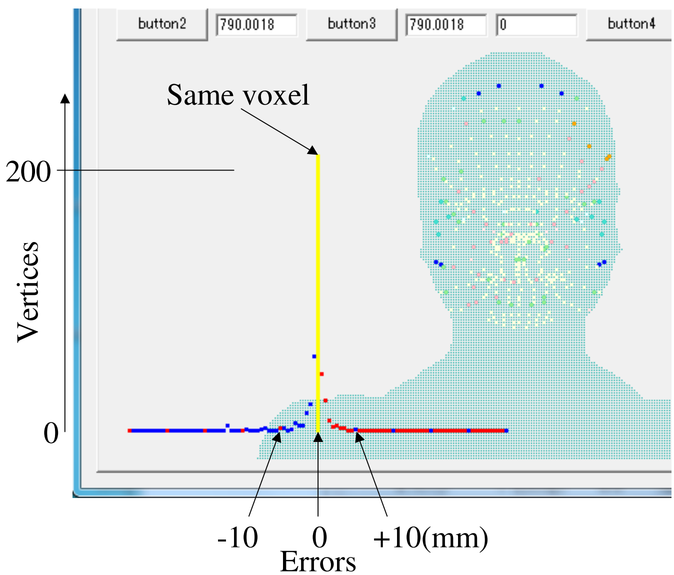


Fig. 7. A histogram for fitting results

Fig. 7 shows a histogram for fitting results.

After some iteration for escaping local minima, the score of total errors for 428 vertices is under 750. (This score has no unit. However in our experiments, the unit of the score is equal to “mm”.) More than 50% of vertices are just on the reference targets (same voxel). More than 85 % of vertices are fitted within +/- 3mm of errors (neighbor voxel), and 95% are within +/- 5mm. It is good results, and we can find leaner silhouette of the face. It is difficult to achieve the score by manual fitting operation with the error display.

The lowest errors value is not necessarily bring the best fitting result. We don't use "asymmetric" parameters for our evaluation like "Face Shear", "Eye Pop", "Crooked Nose", and "Shift Mouth" listed on Table 1. Although the use of these parameters may return the lowest errors value, it often brings an unnatural face with parameter settings that the human never choose.

5 Conclusion

In this paper we have described the empirical study of automatic facial fitting for Second Life avatar.

We used MRI data as our reference model. It is useful for facial modeling. It will be helpful for computer animation field if it becomes easier to make use of raw image data measured by MRI for medical use.

After the iterations, more than 50% of vertices are just on the reference targets, and more than 85 % are within +/- 3mm of errors. It is better than that of manual fitting operation with the error display.

Both MRI data and textures will improve the accuracy of facial fitting.

References

1. Blanz, V., Vetter, T.: A Morphable Model for the Synthesis of 3D Faces. In: Proc. of ACM SIGGRAPH 1999, pp. 187–194 (1999)
2. Blanz, V., Vetter, T.: Face Recognition Based on Fitting a 3D Morphable Model. IEEE Trans. On Pattern Analysis and Machine Intelligence 25(9), 1063–1074 (2003)
3. Dohi, H., Ishizuka, M.: A Life-like Agent Interface System with Second Life Avatars on the OpenSimulator Server. In: Proc. of 13th Int'l Conf. on Human-Computer Interaction, HCII 2009 (2009)
4. Koda, T., Maes, P.: Agents with Faces: The Effects of Personification of Agents. In: Proc. HCI 1996, pp. 98–103. The British HCI Group (1996)
5. Maes, P.: Agents that Reduce Work and Information Overload. Commu. ACM 37(7), 31–40 (1994)
6. Nagaoka, T., Watanabe, S., Sakurai, K., Kunieda, E., Watanabe, S., Taki, M., Yamanaka, Y.: Development of Realistic High-Resolution Whole-Body Voxel Models of Japanese Adult Male and Female of Average Height and Weight, and Application of Models to Radio-Frequency Electromagnetic-Field Dosimetry. Physics in Medicine and Biology 49, 1–15 (2004)
7. OECD Health Data 2010 – Version (October 2010), `http://www.oecd.org/`
8. Patel, A., Smith, W.A.P.: 3D Morphable Face Models Rivisited. In: Proc. of Int'l Conf. On Computer Vision and Pattern Recognition, CVPR 2009, pp. 1327–1334 (2009)
9. Second Life official site, `http://secondlife.com/`
10. Second Life Marketplace, `https://marketplace.secondlife.com/`

Reflected in a Liquid Crystal Display: Personalization and the Use of Avatars in Serious Games

Shan Lakhmani and Clint Bowers

Department of Psychology, University of Central Florida,
Orlando, FL, USA 32816
shan.lakhmani@knights.ucf.edu, bowers@mail.ucf.edu

Abstract. Personalization, in the realm of Serious Games, is the extent to which users believe that the digital environment is tailored to their characteristics and preferences. This belief can have major repercussions for a user's experience with the game and can subsequently be used to maximize the return on investment for serious game designers. Other factors that influence users' personalization in games include how the game affects the users' perception of self, presence in the game, and social relationships developed in the game. Users' avatars influence all of these factors. This goal of this paper is to examine the research done into avatars and personalization and presents it in the context of serious games research.

Keywords: avatars, personalization, presence, immersion, serious games.

1 Introduction

One of the benefits of using games to accomplish goals in addition to entertainment is that the experience can be tailored to the individual user in a way that other means cannot. Developers can personalize the game experience through the use of adaptive content [1] or through the establishment of a relationship between the user and the narrative framework [2]. Not only does this bond yield a motivated user, but also one who more actively cognitively processes the material, and thus finds the material in the game less difficult to understand [2]. The development of a personalized game can yield positive, tangible results from the users [2,3,4,5]; however, what can developers specifically manipulate to make a more personalized game? This text explores the use of player avatars and the effects of design choices, in accordance with existing avatar research.

2 Personalization

Avatars can be a useful means of developing a personalized relationship between a user and a game. There are only so many things one can do with avatars. One can shift perspective between first person and third person perspectives and one can implement a means of avatar customizability. One can either have: no customization by giving users an avatar; limited customization by giving users a choice of pre-made avatars; or full

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 237–242, 2011.

customization in which the user can edit a variety of their avatar's features [5,6]. Advergames, an interactive marketing technique that combines a game with branded material, have used full avatar customization as a means to make the game more personal [3]. Specifically, by giving users the opportunity to make their avatars represent their identity more accurately, advertisers affect the emotional experience the users have, leading to stronger feelings towards the brand [3]. The degree to which a game can positively influence the "affective dimension" of interactive media usage can be understood, and subsequently examined, as immersion [6]. Furthermore, when users are given more choice, and subsequently more control over the virtual environment, they show increased intrinsic motivation [3]. Humans exhibit social reactions to avatars, even if they do not represent actual users [4]. Not only are other players' avatars the means through which users form social bonds [7], but users have been shown to establish parasocial relationships with their own avatars [8].

Thus, avatars are a means of affecting personalization. Personalization affects how much a user gets out of a serious game. Avatars can also affect other factors—users' performance and interpretation of identity, level of immersion, and social relationships—which in turn affect personalization. This paper will explore these factors, how avatars can affect these factors, and what one can do to maximize the deliverables of serious games using these factors and avatars, in accordance with current literature.

2.1 Identity and Personalization

Using an avatar is a means to, quite literally, see oneself in a game. Subsequently, users see the game as a much more personalized experience because they "inhabit" that virtual world. Although, it is not exactly true to say that they, themselves, inhabit that virtual world. People tend to have two sets of schema for themselves; an actual, representative schema and an idealized schema [5,6,7]. When given the opportunity to customize an avatar, users will make a character that more closely resembles their ideal selves than they themselves do [5,6,9]. However, because the avatars that users make will also represent themselves, the avatar ends up being an amalgamation of the ideal and actual selves [7]. In fact, according to Social Cognitive theory, when children are looking for models with whom to identify, they seek models who are like them, thus accessible, but who have more positive characteristics than they have; the avatar, both self-representative and ideal, is the prime focus on which users can identify [5]. Jin [6] found that users who were explicitly asked to create an avatar that reflected their ideal selves reported greater game interactivity, and subsequently greater control over the digital environment, than those who were instructed to create an avatar based on their actual selves. Furthermore, identifying with a subject, also according to Social Cognitive theory, leads to greater learning [5]. So, avatar customization can lead to increased knowledge retention as well as increased feelings of control, which encourages personalization.

Seeing oneself act, even if it is through a digital representation, encourages embodiment, a tendency that can be used by developers to further the goals of their serious game. Avatars are users' virtual bodies, which are thereby cognitively imbued by users with properties of the self [10]. In accordance with existing literature on disembodiment, the same part of the brain used to imbue the action of tools being used as an extension of the self, the right parieto-temporal occipital junction, is also activated when one controls an avatar [10]. This effect appeared even though the image used was selected by the experimenter. However, unlike simpler foci of disembodiment, like a hammer, the avatar often resembles an individual. Through this representation, the user's mental model of the self, and even the user's behavior outside the virtual environment, can be affected [3]. Allowing users to totally customize their own avatars makes the virtual environment seem more real; however, due to this reciprocal relationship between user and avatar, if users are given a choice of pre-made avatars, their ideas of self are actually altered by the characteristics of those avatars [3].

A minor change in identity and subsequent behavior can be seen when avatars are used in a social networking context. Users who were represented by attractive avatars were more willing to approach members of the opposite sex, when compared to those represented with less attractive avatars [9]. Inducing a self-concept discrepancy, by making more salient the gap between the ideal self and the actual self, can lead to a more noticeable change in behavior [6]. Researchers found that users who viewed their avatars running on a treadmill engaged in significantly more exercise behavior in comparison to users who saw their avatars engage in more indolent activities [5].

So, the choices that developers give users in terms of avatars can affect how strongly the user personalizes with the virtual environment as well as how users see themselves. Needless to say, the ability to affect the user's identity can prove invaluable if the serious game developer is trying to effect behavioral change. If learning is the goal, implementing avatar customization can give users an ideal teacher to look up to: themselves.

2.2 Immersion and Personalization

An avatar is the medium through which one can inhabit a virtual world. The experience of actually inhabiting a virtual self—be it a physical representation, as in third-person virtual environment, or a fully psychological representation, as in first-person virtual environments—is described as self-presence [3]. If users feel sufficient self-presence, then they are immersed within their virtual world; if they are sufficiently immersed, they reach a state of flow [11]. To help users reach this flow state, a state where they perform at their optimal level to the exclusion of the outside world and the passage of time, the factors that disturb users' smooth action must be designed out, while factors that allow them to invest themselves must be maximized [12]. Developments that interrupt the experience, such as a poor user interface or an illogical solution to an in-game problem, prevent users from "buying into" the virtual world and prevent flow [12]. If users are provided with adaptive choices, their experience will be much more personalized, and hence be more conducive to eliciting a flow state [13]. So, when the game tailors itself to the user's needs and requirements, the user's immersion into the virtual environment increases. One of the major benefits for immersion is increased time-on-task [14]; as time playing the game increases, so does time exposed to the material in the game.

Avatar customization makes the world more immersive [3] because it allows self-presence to manifest through physical authenticity [15]. A high level of immersion amongst users, in relation to a game, is associated with positive learning outcomes and intrinsic motivation for the activity [3,15]. Therefore, encouraging personalization through the use of avatar customization can lead to a useful ROI for serious game makers.

2.3 Socializing with Avatars

As previously stated, the user can inhabit a virtual world through a mental representation, through a first-person perspective [3]. This design choice can be used to create a compelling game [5] which invokes a user's arousal, presence, and identification; however, when users are given a choice of avatar, they show more arousal, presence, and identification in the third-person perspective than in the first-person perspective [3].

Personalization does not have to be purely about the self. While users can form parasocial relationships with their own avatars, they can also form social bonds with the digital figures that populate a virtual world, be they human-controlled or otherwise. When users see an avatar, be it their own, someone else's, or even one controlled by a computer, they interpret it as they would a human [14]. When playing a game with a digital figure, if the user believes that there is human agency behind the avatar, the anterior paracingulate cortex, the part of the brain known for attributing independent mental states different from our own to others, activates [14]. Moreover, the person perception processes that activate when they see a digital figure are the same as if they saw a videotaped person [16]. When the users believe that they are digitally interacting with a human figure, they show a variety of effects which could be beneficial to the goals of a serious game designer. The development of a relationship with a human-like figure leads to a positive learning experience and increases the users' motivation [16]. It has also led to an increased amount of science information learning in a virtual reality learning experiment [14]. So, allowing users to interact with one another digitally can further the goals of a serious game designer. But, sometimes multiplayer worlds are unfeasible. Will users develop these personalized relationships with computer-controlled agents?

The answer is mixed. If any human-like cues are received from a digital figure, even if there is not human agency behind that cue, people exhibit the same social reactions as they would from human-to-human interaction, with the same parts of the brain activating [16]. People can form parasocial relationships with non-human agents and they subsequently respond as if in a typical relationship [8]. This parasocial relationship can support learning as well; "animated pedagogical agents" can transmit information using nonverbal behavior and thus support student learning [4]. Furthermore, when Moreno's & Mayer's [2] computerized agent used personalized language to aid in teaching users botanical information, users felt more immersed in the virtual environment and they reported that the learning task was easier. So, users respond to computer controlled agents similarly to those controlled by other users and they can benefit from them. However, the use of digital agents has its pitfalls. Users in a virtual reality environment learned less science information from a digital presentation controlled by a computer, than with one controlled by another user [14]. Also, in competitive contexts, people report a greater dislike for computerized agents in comparison to human users [2], and they subsequently feel higher levels of aggression after the game is played [14]. Consequently, heart rate acceleration, a measure of cognitive effort and positive valence, is higher in comparison to baseline when one plays with a user's avatar rather than a computerized agent; moreover, users lose interest in the game if they compete with a computerized agent and are defeated [14]. For a serious game designer, while computerized agents can be useful, they have some definitive drawbacks.

The important aspect of forming the relationship is, of course, the user's belief. The user's feeling of agency over the virtual world, which is the feeling of being able to affect the world through the avatar, is described as self-presence [15]. The user's feeling of social interaction, regardless of technological mediation is described as social-presence [3,15]. The users' sense of interacting with another social being, their social-presence, seems to be the underlying factor behind the increased personalization and subsequent positive results stated above. Users can be persuaded to accept computerized agents as human—thus gaining the benefits from human-human interaction—but creating a context where this acceptance may occur may be difficult [2].

Individual differences also can play a part in creating social relationships with digital figures. People with interdependent self-construal tend to see themselves in terms of relationships. In the virtual environment, these people form social relationships to other people's avatars (seeing them as an extension of a separate person) and they form a parasocial relationship with their own avatar [8]. Because these relationships are emphasized, these users show a higher level of self-presence, with all the benefits that entail [8]. Developers can reap the benefits of this self-construal, however, by priming interdependent self-construal with multiplayer or group player mechanics [8].

The great effect that computer-controlled avatars can have, both positive and negative, emphasizes the need for strong, careful design in serious games. If a designer can create a computerized agent that can mimic a human relationship, the developer of the serious game can gain the benefit of social presence without the drawbacks. This stronger relationship with the virtual world will strengthen the personalization of the world to the user and will contribute towards the goal of the serious game.

3 Conclusion

According to the literature, avatar implementation can affect personalization of the user as well as the factors that feed into personalization. The user's game experience, and what they subsequently receive from the game, can be altered depending on how a designer implements avatar creation in a serious game. A highly customizable avatar can lead to increased immersion and disembodiment, both of which yield certain benefits in terms of user experience. However, a more limited customization or even an avatar choice system requires less effort to implement and can effect behavioral change. Depending on the desired outcome, certain kinds of implementation are more useful than others. However, while some initial recommendations can be made, existing research has only scratched the surface of avatar utility. More research into the specifics of avatar implementation would allow for more specific designer recommendations for avatar implementation, especially in the face of different contexts.

References

1. Peirce, N., Conlan, O., Wade, V.: Adaptive Educational Games: Providing Non-Invasive Personalised Learning Experiences. In: 2nd IEEE International Conference on Digital Games and Intelligent Toys Based Education, pp. 28–35. IEEE Press, New York (2008)
2. Moreno, R., Mayer, R.E.: Personalized Messages That Promote Science Learning in Virtual Environments. J. Edu. Psych. 96, 165–173 (2004)
3. Bailey, R., Wise, K., Bolls, P.: How avatar customizability affects children's arousal and subjective presence during junk food-sponsored online video games. Cyber. Psych. & Beh. 12, 277–283 (2009)
4. Krämer, N.: Psychological research on embodied conversational agents: The case of pedagogical agents. J. Media Psych: Theories, Methods, and Applications 22, 47–51 (2010)
5. Lewis, M.: The effect of learning type and avatar similarity on learning outcomes in educational video games. Dissertations & Theses: Full Text 70 (2010)
6. Jin, S.: Avatars mirroring the actual self versus projecting the ideal self: The effects of self priming on interactivity and immersion in an exergame. Wii. Fit Cyber. Psych. & Beh. 12, 761–765 (2009)
7. Bessière, K., Seay, A., Kiesler, S.: The Ideal elf: Identity Exploration in World of Warcraft. Cyber. Psych. & Beh. 10, 530–535 (2007)
8. Jin, S., Park, N.: Parasocial interaction with my avatar: Effects of interdependent self-construal and the mediating role of self-presence in an avatar-based console game. Wii. Cyber. Psych. & Beh. 12, 723–727 (2009)
9. Vasalou, A., Joinson, A., Bänziger, T., Goldie, P., Pitt, J.: Avatars in social media: Balancing accuracy, playfulness and embodied messages. Int. J. Hum Comp. Stud. 66, 801–811 (2008)
10. Corradi-Dell'Acqua, C., Ueno, K., Ogawa, A., Cheng, K., Rumiati, R.I., Iriki, A.: Effects of shifting perspective of the self: an fMRI study. Neuroimage 40, 1902–1911 (2008)
11. Kenny, R.F., Gunter, G.A.: Endogenous fantasy-based serious games: Intrinsic motivation and learning. Int. J. Soc. Sci. 2, 8–13 (2007)
12. Van Eck, R.: Digital game based learning: It's not just the digital natives who are restless. Educause Review 41, 17–30 (2006)
13. Chen, J.: Flow in Games (and Everything Else). Communications of the ACM 50, 31–34 (2007)
14. Lim, S., Reeves, B.: Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player. Int. J. Hum-Comp. Stud. 68, 57–68 (2010)
15. Feldon, D.F., Kafai, Y.B.: Mixed methods for mixed reality: Understanding users' avatar activities in virtual worlds. Edu. Tech. R&D 56, 575–593 (2008)
16. Krämer, N.C., Bente, G.: Personalizing e-learning. The social effects of pedagogical agents. Edu. Psych. Rev. 22, 71–87 (2010)

Leveraging Unencumbered Full Body Control of Animated Virtual Characters for Game-Based Rehabilitation

Belinda Lange, Evan A. Suma, Brad Newman, Thai Phan, Chien-Yen Chang, Albert Rizzo, and Mark Bolas

Institute for Creative Technologies, University of Southern California
12015 Waterfront Drive, Playa Vista, CA 90294
United States of America
lange@ict.usc.edu

Abstract. The use of commercial video games as rehabilitation tools, such as the Nintendo® Wii Fit™, has recently gained much interest in the physical therapy arena. However, physical rehabilitation requires accurate and appropriate tracking and feedback of performance, often not provided by existing commercial console devices or games. This paper describes the development of an application that leverages recent advances in commercial video game technology to provide full-body control of animated virtual characters with low cost markerless tracking. The aim of this research is to develop and evaluate an interactive game-based rehabilitation tool for balance training of adults with neurological injury. This paper outlines the development and evaluation of a game-based rehabilitation tool using the PrimeSense depth sensing technology, designed to elicit specific therapeutic motions when controlling a virtual avatar in pursuit of in-game goals. A sample of nine adults participated in the initial user testing, providing feedback on the hardware and software prototype.

Keywords: video game, balance, stoke, camera tracking.

1 Introduction

Stroke incidence, new or recurrent, is approximately 800,000 every year, and as the population ages, this number is expected to rise [1]. The neurological impairments that can result from a stroke, especially impairments in postural control, can affect a person's balance and mobility in everyday activities. Improving postural control is one of the main challenges for stroke rehabilitation. To achieve this, individuals are taught to bear weight through their affected lower extremity, resulting in increased balance and gait velocity, decreased fall risk, and greater participation in activities of daily living. Conventional physical therapy techniques typically use visual biofeedback and force plate systems to encourage weight shift onto the impaired side or limb in order to improve weight shift in sitting, standing and during gait [2,3,4]. However, these techniques provide limited objective measurement of performance and typically lack engaging content to motivate individuals during training.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 243–252, 2011.

The use of commercial video games as rehabilitation tools, such as the Nintendo® Wii Fit™, has recently gained much interest in the physical therapy arena. While anecdotal evidence suggests that games have the potential to be powerful motivators for engaging in physical activity, limited published research exists on the feasibility and effectiveness of leveraging the motion sensing capabilities of commercially available gaming systems for rehabilitation [5,6,7,8,9]. Initial case studies have demonstrated that the use of video games has some promise for balance rehabilitation following stroke and spinal cord injury [7,8,9,10]. However, currently available commercial games may not be suitable for the controlled, focused exercise required for therapy. Usability studies have found that some commercially available games provide negative auditory and visual feedback during therapy tasks [6,11]. These observations demonstrate the importance designing games specifically for rehabilitation, a design approach that has been investigated by several recent researchers [11,12,13]. However, the limitations of the commercial video game motion sensing technology have been a challenge to achieving this goal. Motion tracking controllers such as the Nintendo® Wiimote are not sensitive enough to accurately measure performance in all components of balance. Additionally, users can figure out how to "cheat" inaccurate trackers by performing minimal movement (e.g. wrist twisting a Wiimote instead of a full arm swing). Physical rehabilitation requires accurate and appropriate tracking and feedback of performance. As such, we are developing applications that leverage recent advances in commercial video game technology to provide full-body control of animated virtual characters. A key component of our approach is the use of newly available low-cost depth sensing technology from PrimeSense, the company that developed the sensor hardware in the popular Microsoft® Kinect. This technology, along with software from OpenNI, provides markerless full-body tracking on a conventional PC using a single plug-and-play USB sensor. Not only does this approach provide a fully articulated skeleton that describes the user’s body pose, but it does so without encumbering the user with tracking devices or markers. This appears to provide more natural and intuitive interaction, without having to alter natural motor movements to accommodate the tracking hardware. The depth sensing camera allows the user to puppet a virtual character on screen that directly represents their movements and poses in the real world. This system approach enables a game-based rehabilitation tool that is tailored to individual therapy goals. This application is being developed and refined through the process of user-centered design, incorporating feedback from key stakeholders (clinicians, patient groups and care takers) and undergoing iterative feedback and refinement phases.

2 Method

The aim of this research is to develop and assess an interactive game-based rehabilitation tool for balance training of adults with neurological injury. We developed a game-based rehabilitation task designed to elicit specific therapeutic motions when controlling a virtual avatar in pursuit of the in-game goal.

2.1 Hardware

Full-body interaction with the game was provided by the PrimeSense Reference Design, a USB plug-and-play device that uses an IR projector along with standard RGB and infrared CMOS image sensors (Figure 1). To construct a depth map, the sensor uses a proprietary algorithm to resolve the pattern produced by projecting coded infrared light onto the scene geometry. This system has a field-of-view of 58 degrees horizontal and 45 degrees vertical, and generated depth maps with a resolution of 640x480 at 30 frames per second.

Fig. 1. PrimeSense Reference Design USB plug-and-play device

2.2 Software Architecture

The software components of the sensing package are the OpenNI and NITE frameworks, which provide user identification, scene segmentation, and skeleton tracking. To create the game, we used the Unity3 game engine along with an OpenNI wrapper developed by PrimeSense to integrate these frameworks with the game engine. The engine provides a C# API and flexible editor which allows for rapid development of games. This rapid development cycle enables faster iteration of game mechanics and ultimately results in more specificity for tailoring games to address patient's individual disabilities.

2.3 Game Overview

The game world involves a precious jewel mine where the player assumes the role of a miner who rides a railroad cart down a mine shaft and gathers jewels from the shaft walls. The shaft is uniformly cylindrical with eight jewels arranged in a ring with the player's avatar centered in the middle of the screen (Figure 2). In order for the player to successfully gather all the jewels they must reach out from the center of the screen and touch each jewel individually with their hand.

Fig. 2. The game is set in a jewel mine. The player's character is situated in the center of the mine shaft with eight gems placed around them in a ring.

2.4 Calibration

The game begins with a calibration step which maps the range of motion of the player's upper limbs. After striking a default calibration pose for a few seconds (Figure 3) the player's skeleton is captured and tracked without any attachment of devices or markers.

Fig. 3. User calibration step: The player must strike a default pose and the player's upper torso, arms, hands and head are represented on the screen

The supporting OpenNI software recognizes when a human figure enters the frame and sends joint position data from the PrimeSense camera to the game. This position data is then transformed into joint orientations with simple trigonometry algorithms. The player's upper torso, arms, hands and head are represented on-screen by an abstract avatar made up of white cylinders for the arms, chest and hands and a sphere for the head (Figure 3). The length of the cylinders for the chest and arms are scaled to match each player's specific anatomy.

To map the individual player's range of motion, the player moves their arms as far as possible in 3D space forward and outward in an arc over eight radiating lines. This step should be guided by a clinician to encourage movements that are appropriate for the player and specific therapy goals. As the player holds their hand at the intersecting point of each line a jewel is placed at the point of intersection (Figure 4). Upon completing the placement of each jewel, the positions are saved and loaded into the game itself.

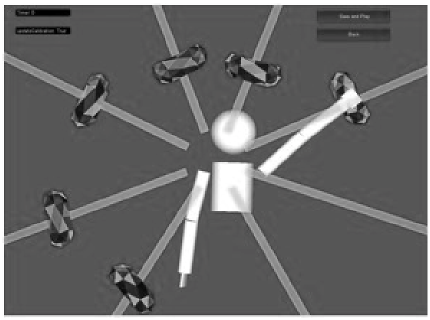


Fig. 4. Calibration step: to map the individual player's range of motion, the player moves their arms as far as possible outward in an arc over eight radiating lines

2.5 Gameplay

In the game, all eight jewels are present and visible in a series of rings that extend as the player moves through the mine shaft. Individual jewels glow to indicate when they can be gathered. The order in which the jewels glow is controlled by three different pre-defined patterns that the clinician can select before a session begins: Sequential pattern, Simon pattern or Sam pattern. During the Sequential pattern, jewels light up one by one. The player's goal is to collect the gems as they light up. During the Simon pattern, the jewels light up in the classic "Simon" game pattern and the player must remember the sequence in which the jewels light up and touch them in that order. The Sam pattern was designed to be a modified "Simon" pattern where the recall task is removed and the jewels remain lit until the player collects them. The Sequential and Sam patterns each have two difficulty settings.

2.6 Avatar Representation

Before starting the game the clinician is given the option of displaying the player's avatar as either a full torso with arms and head or simply displaying the hands (Figures 5a and 5b). This avatar representation allows the clinician to control how much feedback the player receives about their body position during the game. The full upper torso is tracked within both representation settings, however the hands only representation provides visual representation of the hand points of the skeleton.

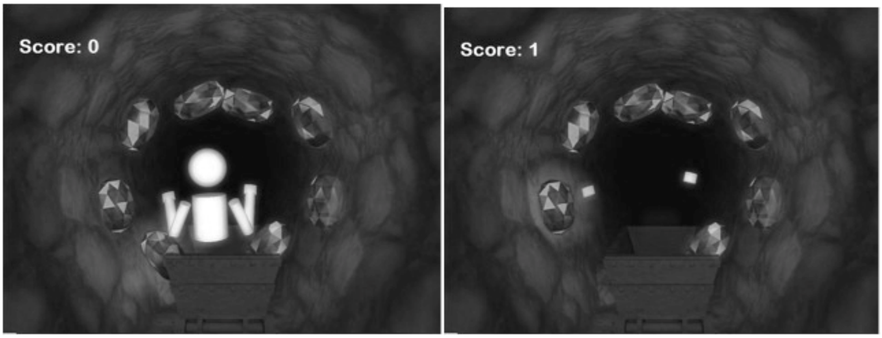
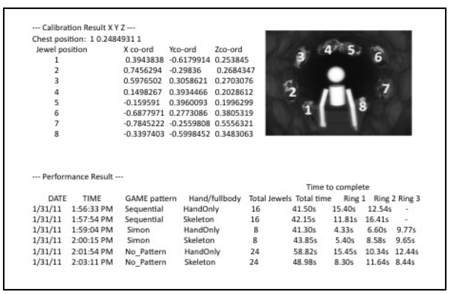


Fig. 5. Avatar representation: The game can be played with a upper torso skeleton avatar (a) or with only the player's hands represented on the screen as two squares (b)

2.7 Data Recording

The player's range of motion calibration settings are saved as a player profile and can be loaded in subsequent sessions. Additionally, the duration of time it takes patients to complete each jewel ring task and the total time it takes to complete the game are recorded and reported (Figure 6).



--- Calibration Result X Y Z ---

Chest position: 1 0.2484931 1

Jewel position	X co-ord	Yco-ord	Zco-ord
1	0.3943838	-0.6179914	0.253845
2	0.7456294	-0.29836	0.2684347
3	0.5976502	0.3058621	0.2703076
4	0.1498267	0.3934466	0.2028612
5	-0.159591	0.3960093	0.1996299
6	-0.6877971	0.2773086	0.3805319
7	-0.7845222	-0.2559808	0.5556321
8	-0.3397403	-0.5998452	0.3483063

--- Performance Result ---

					Time to complete			
DATE	TIME	GAME pattern	Hand/fullbody	Total Jewels	Total time	Ring 1	Ring 2	Ring 3
1/31/11	1:56:33 PM	Sequential	HandOnly	16	41.50s	15.40s	12.54s	-
1/31/11	1:57:54 PM	Sequential	Skeleton	16	42.15s	11.81s	16.41s	-
1/31/11	1:59:04 PM	Simon	HandOnly	8	41.30s	4.33s	6.60s	9.77s
1/31/11	2:00:15 PM	Simon	Skeleton	8	43.85s	5.40s	8.58s	9.65s
1/31/11	2:01:54 PM	No_Pattern	HandOnly	24	58.82s	15.45s	10.34s	12.44s
1/31/11	2:03:11 PM	No_Pattern	Skeleton	24	48.98s	8.30s	11.64s	8.44s

Fig. 6. Data recorded from game includes player profile and time to complete the in-game tasks

2.8 User Testing

To evaluate the usability of the rehabilitation game, participants receiving balance training following stroke were recruited from two Rehabilitation Clinics in the Los Angeles area. Participants were asked to perform two game-based tasks, which were randomized for each participant. The two game-based tasks differed in presentation of participant's character on the screen: upper torso skeleton and hands only view. The Sequential pattern was used for both tasks. The researchers observed participants playing the game, providing assistance when needed. Following the interaction, the participant was asked to complete a structured interview about their experience. Clinicians attending the user testing session with a participant were also asked to provide feedback. Game performance, researcher observations and player feedback were analyzed to investigate how these game design criteria influence the performance of tasks that are relevant for rehabilitation.

3 Results

To date, a sample of nine participants (four females and five males), aged 52-78 years have been recruited in the study. Participants, receiving therapy following stroke (six months to seven years post stroke), were able to sit or stand independently or with light assistance. Three participants had previous experience playing video games. Of those three participants, only one owned and used a Nintendo® Wii™ and Nintendo® Wii Fit™ in their home.

Four participants played the game in their wheelchair, while five participants played the game in standing. In standing, three participants required stand-by assistance and two participants required light-moderate assistance.

3.1 Calibration

All of the participants had difficulty performing the calibration pose. Two of the nine participants could not complete the calibration step, and therefore were unable to play the game and could not complete the rest of the testing session. Limitations in range of motion of the impaired upper limb prevented these participants from being able to perform and maintain the required position. The final seven participants required assistance from a clinician to provide active assisted or passive shoulder flexion, abduction and external rotation of the impaired upper limb in order to perform and maintain the calibration pose.

3.2 Game Play Observations

Participants, both in a wheelchair and in standing, were able to complete the game-based tasks with assistance where required. During game play, participants tended to need more instruction initially at the start of the game. However, following the completion of one to three sets of the jewels, participants appeared to gain sufficient understanding of the task to perform without instructions. Five of the participants had particular difficulty with perception of the depth of the jewels within the 3D space.

3.3 User Feedback

The structured interview questions provided qualitative data about user perceptions of the technology, motivation to use the technology and specific feedback on potential changes to improve the existing system. Participants provided the research team with suggestions for improving the instructions for the game and game play elements (such as scoring and sound effects).

Overall, participants reported the game to be challenging and fun. Specific comments related to level of challenge and enjoyment were:

"I think it was more exciting than working muscles just in the gym"
"You really made me work hard!"

"I enjoyed it but it made me work really hard to get all those gems"
"I felt like I was really slow at first but once I got the hang of it I had fun"
"It was good. I want to play it again."
"When I was collecting the gems I wasn't thinking about how hard I was working"

Each of the seven participants chose the skeleton avatar in preference over the hands-only avatar. Participants stated that seeing the full upper body avatar gave them more information about how they were moving and helped them to collect the jewels.

When asked to describe how the game compares to current therapy activities, four participants responded that the game was similar to activities they are currently asked to perform in therapy sessions. Five out of the seven participants that were able to play the game stated that they would like to play the game again and could see themselves playing the game as part of their regular therapy once it is completed.

3.4 Clinician Feedback

A sample of four clinicians consented to provide feedback on the system. Overall, the clinicians stated they were excited about the use of this type of technology within the clinical setting. One clinician stated he had been investigating commercially available games for use with his patients but had found many of the game-based tasks to be too difficult. He stated that the concept of calibrating the system to an individual user and having control over the game-based task was an important feature in the current prototype. Furthermore, the option to change the avatar view from skeleton to hand only views, providing less feedback, therefore increasing the level of challenge for the patient appealed to the clinicians.

4 Discussion

The release of the PrimeSense 3D depth sensing technology provides a low-cost option to capture users' full body movement in 3D space for interaction within game activities without the need for the user to hold an interface device. Whilst this provides developers, researchers and clinicians with access to low-cost tracking technologies that can be more easily implemented within tailor made software programs, accessibility of the system is still limited to people who can actively or passively perform the calibration pose. At this point in time, people with disabilities that limit them from being able to perform the calibration pose, cannot interact with the system.

Calibration is an important piece of tailoring game-based applications for individuals with varying levels of ability. The current OpenNI software technology will track the user only once a calibration pose has been performed. The calibration pose requires the user to stand or sit facing the camera, place the arms out to the side with the shoulder in external rotation and at 90 degrees abduction and the elbow at 90 degrees of flexion (Figure 2a). The importance of user-centered design is highlighted when issues such as calibration are exposed earlier in the development process to allow for changes that will make the system more accessible to the key stakeholders that will be using the system.

Despite difficulties with calibration, initial user feedback supported the use of game-based rehabilitation and provided a range of suggestions for system improvement. The suggestions and comments are currently being themed and will be discussed, evaluated and prioritized by the research and development team.

Feedback from clinicians was supportive of the development of a flexible system that provides the option to tailor the game-based task to the user. Clinicians liked the option to change the avatar view from skeleton to hand only views. Changing the appearance of the avatar, whilst still tracking the skeleton, allows the clinician to remove visual feedback provided to the patient. Removing this external feedback, encourages the patient to rely on internal sensory-perceptual information to control their movement [14]. Providing less visual feedback about the player's movement increases the level of difficulty and this was supported in the participant's unanimous choice of the skeleton avatar as the easiest to understand and use. However, the use of the hands only avatar representation is important progression in therapy guided by motor learning principles [14].

The research team have been developing a Flexible Action and Articulated Skeleton Toolkit (FAAST) with a goal to make a general purpose software environment that enables many applications to quickly be modified to use depth sensing technolgoies (http://projects.ict.usc.edu/mxr/faast/).

Some of our early work and updated videos can be viewed at

http://www.youtube.com/user/AlbertSkipRizzo#p/u/5/geyIvG4uKxY

5 Conclusions

This preliminary research supports the development of a novel and flexible game-based rehabilitation system for balance training. The initial user testing provided insight into key features that need modification or improvement to develop an accessible and user friendly rehabilitation tool with an emphasis on custom calibration and kinematic modeling.

Acknowledgments. The authors would like to thank the PrimeSense Corporation and OpenNI community. The authors would also like to thank the clinicians and participants at Precision Rehabilitation and Rancho Los Amigos National Rehabilitation Center for their valuable time and feedback.

References

1. American Heart Association. Heart disease and stroke statistics 2010 update: A report from the American Heart Association. Circulation, 121, e46–e215 (2010)
2. Geiger, R.A., Allen, J.B., O'Keefe, J.: Balance and Mobility Following Stroke: Effects of Physical Therapy Interventions With and Without Biofeedback/Forceplate Training. Phys. Ther. 81(4), 995–1005 (2001)
3. Walker, C., Brouwer, B.J., Culham, E.G.: Use of Visual Feedback in Retraining Balance Following Acute Stroke. Phys. Ther. 80(9), 886–895 (2000)

4. Yavuzer, G., Eser, F., Karakus, D., Karaoglan, B., Stam, H.: The Effects of Balance Training on Gait Late After Stroke: A Randomized Controlled Trial. Clin. Rehabil. 20(11), 960–969 (2006)
5. Gargin, K., Pizzi, L.: Wii-HAB: Using the Wii Video Game System As An Occupational Therapy Intervention With Patients in the Hospital Setting. Health Policy Newsletter (March 2010)
6. Lange, B.S., Flynn, S.M., Rizzo, A.A.: Initial Usability Assessment of Off-the-shelf Video Game Consoles for Clinical Game-based Motor Rehabilitation. Phys. Ther. Rev. 14(5), 355–363 (2009)
7. Deutsch, J.E., Robbins, D., Morrison, J., Bowlby, P.G.: Wii-based Compared to Standard of Care Balance and Mobility Rehabilitation for Two Individuals Post-stroke. In: Virtual Rehabilitation International Conference, Hafia, Israel, June 29- July 2 (2009)
8. Sugarman, H., Weisel-Eichler, A., Burstin, 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, Hafia, Israel, June 29- July 2 (2009)
9. 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, pp.1–6 (2009)
10. Flynn, S.M., Palma, P., Bender, A.: Feasibility of Using the Sony Playstation 2 Gaming Platform for An Individual Poststroke: A Case Report. Journal of Neurologic Physical Therapy 31(4), 180–189 (2007)
11. Lange, B.S., Flynn, S.M., Chang, C., Ahmed, A., Geng, Y., Utsav, K., Xu, M., Seok, D., Cheng, S., Rizzo, A.A.: Development of An Interactive Rehabilitation Game Using the Nintendo® Wii Fit™ Balance Board for People With Neurological Injury. In: Proceedings of ICDVRAT, Chile, August 31–September 2 (2010)
12. Gil-Gómez, J.A., Lozano, J.A., Alcañiz, M.: Nintendo Wii Balance Board for Balance Disorders. In: Virtual Rehabilitation International Conference, Hafia, Israel, June 29– July 2 (2009)
13. Billis, A.S., Konstantinidis, E.I., Mouzakidis, C., Tsolaki, M.N., Pappas, C., Bamidis, P.D.: A Game-like Interface For Training Senior's Dynamic Balance and Coordination. Medicon IFMBE Proceedings 29, 691–694 (2010)
14. Van Vliet, P.M., Wulf, G.: Extrinsic Feedback for Motor Learning After Stroke: What is the Evidence? Disability and Rehabilitation 28(13-14), 831–840 (2006)

Interactive Exhibition with Ambience Using Video Avatar and Animation on Huge Screen

Hasup Lee[1], Yoshisuke Tateyama[1], Tetsuro Ogi[1],
Teiichi Nishioka[2], Takuro Kayahara[3], and Kenichi Shinoda[4]

[1] Graduate School of System Design and Management, Keio University, Japan
[2] Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan
[3] School of Project Design, Miyagi University, Japan
[4] Department of Anthropology, National Museum of Nature and Science, Japan
{hasups,tateyama,ogi}@sdm.keio.ac.jp,
nishioka@slis.tsukuba.ac.jp, kayahara@myu.ac.jp,
shinoda@kahaku.go.jp

Abstract. In this paper, we develop an interactive exhibition system using a video avatar and an animation on huge screen. As the video avatar, we extract background images from recorded video stream using Chroma key and send the stream to remote using TCP/UDP protocols. The animation on huge screen provides immersion and ambience of the ages of displays. We restore clothes, ceremony, crowd, etc. using computer animation. 4K resolution projector with 300 inch screen is used in our system and it makes viewers to feel the ambience of the environment when and where the displays existed.

Keywords: Video Avatar, Museum Digitalization, Interactive Exhibition, Digital Restoration, Huge Screen.

1 Introduction

There have been existed a lot of museums in history. The museum named Museion at Alexandria, Egypt BC300 is considered the first one [1]. They spread all over the world geometrically and there are several museums in most cities. Also their themes are various like history, foods, clothes, etc.

As computer technologies have been developed, many things in our life are digitalized. The museum is researched using these technologies, too. Users can watch the exhibits on the homepage via computer networks. The exhibits are digitized as graphical models [2]. The museum is constructed in virtual world using VR technique [3]. The information of the art is retrieved using computer vision methods [4]. Tourists of museum are guided using digital devices [5]. The museum is one of most digitalized facility.

The exhibition of the museum becomes more interactive if an export explains in front of the viewers. This environment can be developed using the video avatar [6].

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 253–259, 2011.

The applications of the video avatar are a remote experiment with sharing data, a conference system, etc. And the viewers can feel more immersive using huge screen that the animation of exhibits projected into.

In this research, we focus on the interaction and the immersion of viewers. We develop a video avatar for the interactive exhibition system. We extract background images from the video stream using Chroma key and send the stream to remote host via computer networks. Finally, the video avatar is superimposed on the animation on the screen in front of the viewers. For viewers' immersion, the animation on huge screen is used and provides the ambience of the ages of displays. The clothes, ceremony, crowd, etc. of the ages of displays are restored using computer animation. We use 4K resolution projector and 300 inch screen in our system and it provides the feeling of the ambience of the environment when and where the displays existed.

2 Digital Museum

In this research, we restore an animation of the exhibit's age from the old scroll. Then this animation is projected onto a huge screen for viewers to feel immersion. The video avatar is added on this animation for interactivity. The demonstration of whole system is shown in (Figure 1). The animation is restored using described images on the scrolls for viewer to feel the ambience of its age.

Fig. 1. Demonstration of digital museum

3 Video Avatar

Video avatar is developed to share a virtual space between remote places [6]. Video avatar helps the users to point out the 3D data more precisely. It uses stereoscopic devices for input. In our research, we show video avatar on the restoring animation to interact with viewers.

The expert explains about the animation in front of the camera with microphone. The record module records this video stream and the send module transfers this stream to remote host of the museum via TCP or UDP protocols. The record module and the send module communicate with each other using shared memory. The video stream is received by the receive module and projected on the screen by render module using OpenCABIN library. The receive module and the render module share data through shared memory, too.

The viewers watch the animation and listen to the explanation. They can ask questions or comments to the expert with microphone. The sound data stream which handled by sound program is full-duplex like a telephone. The exhibition is interactive using this video avatar system on the exhibits or the animation. The system architecture of video avatar is shown in (Figure 2).

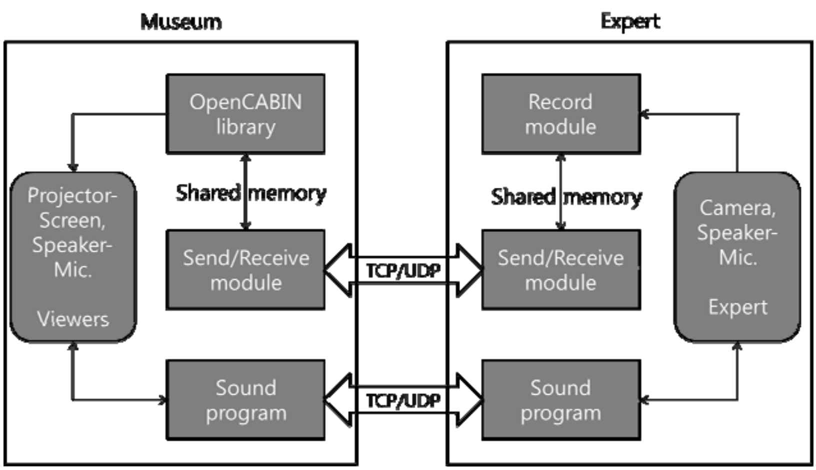


Fig. 2. System architecture of video avatar

4 Animation Restoration

We restore a character animation of an old scroll contain landscape of ceremony. 3D graphical models are created from real photographs of the buildings. We extract their textures from the photographs. The textures of character's clothes, hats, bags, knifes, shoes, etc. are made by referencing exhibits and materials of museums. The example of real image and model of our system are shown in (figure 3).

First, 2D scraps of contour of people of the old scroll are placed. Then we replace them with the 3D characters and animate them. We get motion capture data from the actors for character animation. The voices of characters are dubbed by the professional voice actors and the special sound effect is added for feeling the mood of crowded environment.

Fig. 3. Photograph (left) and model (right) of Kumamoto castle

5 Implementation

We use Sony® digital camera for capturing the expert and its resolution is 800 x 600 pixels. Chroma key technique is used for background removal. The each difference of color values in RGB color space is summed up and the alpha value for pixels is determined by threshold. The expert's video avatar is superimposed on the animation on the screen.

Fig. 4. Send and receive part of video avatar

The record module is implemented using Intel® Integrated Performance Primitives library. The record and the send module communicate using shared memory. The send module sends video stream to receive module via Internet using UDP protocols. The render module accesses the video stream that the receive module received and stored into shared memory. The render module is developed using OpenCABIN library which displays graphical contents on the screens and handles user's inputs like position, direction, joystick values, etc. [7]. The final animation is displayed on 300 inch screen using 4K resolution Sony® projectors. The recording part configuration of video avatar system and the video avatar on the animation are shown in (figure 4).

Fig. 5. Old scroll of Kumamoto Castle

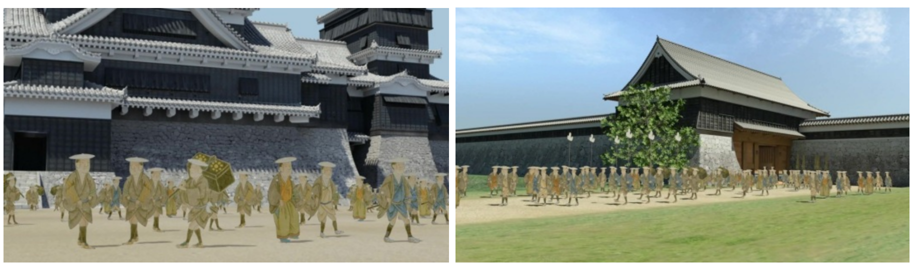


Fig. 6. Character placement of crowd (left) and ceremony (right)

Fig. 7. Character animation of crowd (left) and ceremony (right)

We made the animation of crowd and ceremony of old ages of Kumamoto Castle in Japan. It is restored using old scroll which contains various people and snapshots of ceremony. The scroll is shown in (figure 5). We make the 3D model of castle using texture from real image of it. The 3D characters are created with referencing to other exhibits like clothes, knives, hats, bags, etc.

We scrap people from the scroll and place on our castle model. The left image of (figure 6) is scrap version of the landscape of crowd and the right one is of the old ceremony. After placing scraps, we replace them with 3D characters. The final results are shown in (figure 7).

We get motion capture data from the actors to animate characters. The snapshot of the actor in motion is shown in the left image of (figure 8). The voices of characters are dubbed by the professional voice actors and the special sound effect is added for feeling the mood of crowded environment. Voice actors are dubbing the animation like the right image of (figure 8).

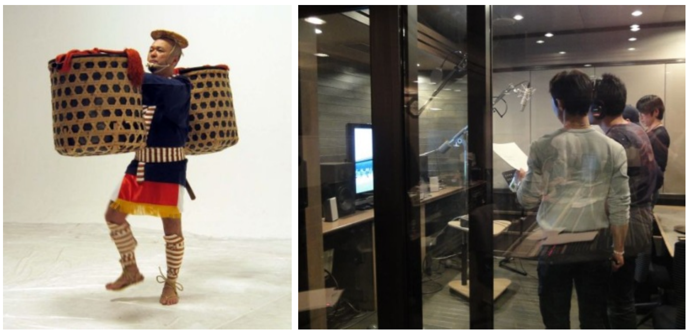


Fig. 8. Recording motion data and voice dubbing

6 Conclusion

We develop a video avatar system for the interactive exhibition. We remove background from the video stream using Chroma key technique and send it to remote via Internet. The video avatar is superimposed on the animation that is restored using old scroll, clothes, knives, hats, bags, etc. First, we place the scraps of people's contour from scroll on the 3d model of environment. Then we replace the scraps with 3D characters and animate them. We use 4K resolution projector and 300 inch screen in our system and it provides the feeling of the ambience of the environment when and where the displays existed. In this research, the interactive, immersive and ambience-feeling exhibition can be developed using the video avatar system, huge screen and the animation restoration.

Acknowledgement. This study was funded partly by "Digital Museum" project of the Ministry of Education, Culture, Sports, Science and Technology, and was supported by Keio University Global COE (Center of education and research of symbiotic, safe and secure system design) Program. The authors would like to thank Shuji Takahashi (The Institute of Cultural Communications, LTD.), Naomi Kudo (TBS Television, Inc.), Kazunori Yoshino (NTT Learning Systems Corporation), Kota Saito and Motonari Yokota (SPIN Inc.) for their supports.

References

1. Doosan Corp. Museum section", EnCyber Encyclopedia, http://www.encyber.com/search_w/ctdetail.php?pno=&masterno=69473&contentno=726156
2. Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J., Fulk, D.: The digital Michelangelo project: 3D scanning of large statues. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques (SIGGRAPH 2000), pp. 131–144. ACM Press/Addison-Wesley Publishing Co. (2000)
3. Pan, Z., Chen, W., Zhang, M., Liu, J., Wu, G.: Virtual Reality in the Digital Olympic Museum. IEEE Computer Graphics and Applications 29(5), 91–95 (2009)
4. Lewis, P.H., Martinez, K., Abas, F.S., Fauzi, M.F.A., Chan, S.C.Y., Addis, M.J., Boniface, M.J., Grimwood, P., Stevenson, A., Lahanier, C., Stevenson, J.: An integrated content and metadata based retrieval system for art. IEEE Transactions on Image Processing 13(3), 302–313 (2004)
5. Abowd, G.D., Atkeson, C.G., Hong, J., Long, S., Kooper, R., Pinkerton, M.: Cyberguide: a mobile context-aware tour guide. Wirel. Netw. 3(5), 421–433 (1997)
6. Tamagawa, K., Yamada, T., Ogi, T., Hirose, M.: Developing a 2.5-D Video Avatar. IEEE Signal Processing Magazine 18(3), 35–42 (2001)
7. Tateyama, Y., Oonuki, S., Sato, S., Ogi, T.: K-Cave demonstration: Seismic information visualization system using the OpenCABIN library. In: Proceedings of the ICAT 2008, pp. 363–364 (2008)

Realistic Facial Animation by Automatic Individual Head Modeling and Facial Muscle Adjustment

Akinobu Maejima, Hiroyuki Kubo, and Shigeo Morishima

Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo Japan
a.maejima@kurenai.waseda.jp, hkubo@suou.waseda.jp,
shigeo@waseda.jp

Abstract. We propose a technique for automatically generating a realistic facial animation with precise individual facial geometry and characteristic facial expressions. Our method is divided into two key methods: the head modeling process automatically generates a whole head model only from facial range scan data, the facial animation setup process automatically generates key shapes which represent individual facial expressions based on physics-based facial muscle simulation with an individual muscle layout estimated from facial expression videos. Facial animations considering individual characteristics can be synthesized using the generated head model and key shapes. Experimental results show that the proposed method can generate facial animations where 84% of subjects can identify themselves. Therefore, we conclude that our head modeling techniques are effective to entertainment system like a Future Cast.

Keywords: Individual Head Model, Automatic Mesh Completion, Facial Muscle Layout Estimation, Key Shape Generation, Facial Animation Synthesis.

1 Introduction

Generating a realistic human head model is still an important theme for computer graphics application such as film and game productions. Many techniques have already been proposed for creating a photorealistic human face [6, 11, 7, 4, 1]. In the latest works of Oleg et al [1], the Digital Emily Project, they have succeeded to synthesize a digital clone which is mistaken to be a real human. However, they have spent approximately 7 months from facial geometry/performance capture to final composition and expensive devices to create it.

In contrast, we have developed the Future Cast System (FCS) which can instantly generate subjects' face models and the film that the generated face models are embedded into a pre-created film using a low cost facial range scanner. The generated facial animations are not so real compared with Oleg's work, but can be recognized by original subjects. In fact, however, 35% of subjects still couldn't identify their character in the movie in which they were appeared. This is because the generated characters lack of subjects' "Personal Characteristics" about whole head geometry with hair and facial expressions [9]. Thus, we have to improve these representations so that the generated character becomes more realistic. So, we focus on how to

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 260–269, 2011.

instantly generate a realistic individual head model which can express person-specific facial expressions from simple and low-cost measurements.

To create a realistic individual head model, we generally fit a template head model to a range scanned whole head geometry using a 3D modeling software by hand. However, in most case, accurate skin-head geometry can't be measured due to occlusions by hair. The result of these regions' scans would be large holes and we need to fit a template model to those parts predictively. Moreover, a whole head range scanner is almost very expensive for practical use.

We propose an automatic head modeling technique with complementing the geometry that it is difficult to measure using template mesh geometry. We employ the low-cost facial range scanner [9] to acquire individual facial geometry. A whole head model is automatically generated based on the automatic marker placement and the template mesh fitting based on energy minimization only from a facial range scan.

We also propose a facial animation synthesis framework based on the facial muscle adjustment. We focus on how to capture and model individual characteristics of facial expressions with simple measurements, and how to synthesize facial animation in real-time. To solve these problems, we firstly estimate an individual muscle layout from subject's videos with facial expressions and then we create individual key shapes using physics-based simulation according to the estimated muscle layout. Facial animations considering personal characteristics are synthesized by blending above individual key shapes. As a result, we can synthesize a facial animation with a person-specific facial expression.

Through the subjective evaluation, we demonstrate the effectiveness of our techniques. Experimental result shows that we archive to synthesize a CG character that 86% of subjects can identify themselves using our techniques. Therefore, we conclude that our head modeling techniques are effective to entertainment systems like a Future Cast.

2 Related Work

Many 3D human head modeling techniques have been developed by [15, 5, 11, 7]. These methods can create a photorealistic animate-able head model by fitting a template head model to input range scan data or photographs. However, manual procedures such as specifying facial landmarks are still required. In this paper, we introduce an automatic facial feature point detector [9] to automatically decide facial landmarks.

Generating a complete mesh from an incomplete mesh with missing parts or noise is known as mesh completion problem, many researchers have been proposed methods to solve this problem [7, 2, 8]. These methods can fill-in missing parts of a scanned geometry using template mesh's geometry. However, correspondences between a small set of vertices on the template mesh and markers on the scanned geometry need to be specified by a user. In our mesh fitting, these marker-vertex correspondences are automatically decided based on fitting the template mesh to a facial texture which has a pixel-to-pixel correspondence for a facial texture [9].

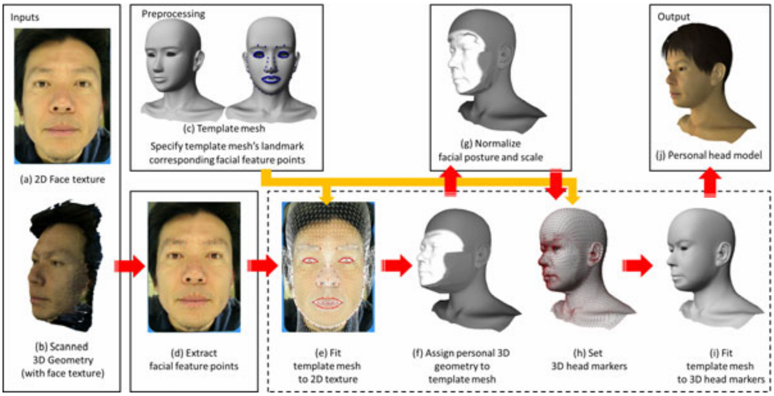


Fig. 1. The overview of our head modeling process

Our mesh fitting approach is inspired from Allen's template mesh fitting based on the energy minimization framework for the registration of body range scans [2]. However, unlike Allen's approach, the marker-vertex correspondences are automatically decided as mentioned above. Moreover, our approach is fast compared with the previous approach because no multi-resolution optimization is required, and can also control the trade-off between the fitting accuracy and the smoothness at each local domain of a template mesh.

Current major 3D facial animation synthesis techniques include: a physics-based approaches developed by [15, 3, 13] and a blend-shape-based approaches where facial expressions are synthesized using linear combinations of several basic sets of pre-created faces for blending developed by [12, 7, 14]. These techniques are able to synthesize highly realistic human facial expressions. Especially in the physics-based approach [13], it can synthesize a person-specific facial expression by modeling a facial anatomical structure for an individual and adjusting the facial muscles layout and the magnitude of the contraction of an each facial muscle. However, it requires the considerably computational time for the simulation. If there are many characters in one frame of the movie, it is difficult to synthesize facial expressions in real-time. Also, blend-shape-based approaches have an advantage that the computation time is clearly less than physics-based approaches. Consequently, we adopt the hybrid-approach that individualized key shapes are generated by the physics-based simulation in offline and the facial animation is synthesized by blending the generated key shapes in real-time.

3 Automatic Head Modeling

3.1 Setting Markers onto Scanned Geometry

Fig. 1 shows the overview of our head modeling process. The range scan data that we deal with in this paper has pixel-to-pixel correspondences between texture and depth map (Fig. 1.(a) and Fig. 1.(b)). To automatically specify markers on an individual scanned geometry, we firstly have to fit the template mesh to 2D face image, and then to acquire the each 3D coordinate corresponding to vertex on the template mesh from the depth map. In advance, we specify 149 landmark points on the template mesh semantically corresponding to facial feature points by hand (Fig. 1.(c)).

In the actual runtime processing, the facial feature point detector [9] automatically detects 149 facial feature points including the outline of the eyes, eyebrows, nose, lips and the facial contour from the face texture at first.

Second, Universal Kriging (UK) is applied to map all of the template mesh vertices onto the 2D face image using the correspondence between detected facial feature points (Fig. 1.(d)) and pre-labeled vertices on the template mesh (Fig. 1.(c)). UK is often applied an oil deposit exploration in Geostatistics and the motion interpolation [10] in computer graphics. Compared with RBFs, UK can predict continuous and much smoother scalar distribution from given scattered samples. We therefore introduce UK to the fitting problem. Let vertex positions of the template mesh corresponding to facial feature points are $\mathbf{C} = \{\mathbf{c}_1,..,\mathbf{c}_N\}, \mathbf{c}_i = (x_i, y_i, z_i) \in \Re^3$, uv coordinates of facial feature points are $\mathbf{u} = (u_1,...,u_N), \mathbf{v} = (u_1,...,u_N) \in \Re$. Where, N is the number of facial feature points, $N = 149$. UK firstly estimates the trend surface $m(\mathbf{c})$ approximating global distribution of $\mathbf{C}$, then calculates the predictive residual by linear combination of residuals between sample data and the estimated trend surface $r_i^x = x_i - m(\mathbf{c}_i)$ and the kernel function $\sum_{i=1}^N b_i^x(\mathbf{c}) r_i^x$. The kernel function $b(\mathbf{c})$ is calculated using a vertex position of the template mesh $\mathbf{c} = (x, y, z) \in \Re^3$ and the variogram function $\gamma^* = (h; \theta)$ (see more detail of the kernel optimization process on [10]). After the kernel computation, we can solve equation (1) to obtain uv coordinates of a vertex position $\mathbf{c}$ of the template mesh (Fig. 1.(e)).

$$u'(\mathbf{c}) = m(\mathbf{c}) + \sum_{i=1}^{N} b_i^u(\mathbf{c}) r_i^u \quad v'(\mathbf{c}) = m(\mathbf{c}) + \sum_{i=1}^{N} b_i^v(\mathbf{c}) r_i^v \tag{1}$$

Third, vertices in the face region of the resulting mesh are replaced by vertices of an individual geometry (Fig. 1.(f)). At this time, the face region's vertices are normalized to match its orientation and scale to the template mesh using the facial pose normalization technique [9] (Fig. 1.(g)). The vertex positions of the resulting mesh are used as markers in the next section (Fig. 1.(h)). However, of course, there would be a boundary between the face region and the other ones if both geometries are quite different. To generate a complete head model, we not only have to realize a seamless connection of both regions' geometries, but also need to preserve accurate individual face geometry as much as possible. To solve this problem, we introduce a new template mesh fitting technique described in the next section.

3.2 Fitting Template Mesh Using Optimized Local Affine Transformation

We describe how to fit the template mesh to markers placed on the scanned geometry using optimized local affine transformation. We assume that a surface on the face at local domain is rigid and estimate affine transformations which map each vertex position on the template mesh to its corresponding marker position locally as-rigid-as possible by solving energy minimization problem.

We define the energy function E as the equation (2). The first term indicates that the position of each vertex of the template mesh should be equal to its corresponding marker position. The second term indicates that the affine transformation of adjacent vertices should be equaled. This term is minimized when change in deformation across the surface, and not the surface itself, is smooth.

$$E = \alpha \sum_{i=1}^{N} w_i^m \left\| \mathbf{T}_i \mathbf{v}_i - \mathbf{m}_i \right\|^2 + \beta \sum_{\{i,j|\mathbf{v}_i,\mathbf{v}_j \in edge(mesh)\}} w_i^s \left\| \mathbf{T}_i - \mathbf{T}_j \right\|_F^2 \tag{2}$$

Where, $\mathbf{v}_i$ and $\mathbf{m}_i$ are i-th vertex and its corresponding marker respectively. $\mathbf{T}_i$ represents the affine matrix of the i-th vertex on the template mesh. $\|\cdot\|_F$ indicates the Frobenius norm which is the sum of the square root of each matrix element. w_i^m and w_i^s are weights to control the fitting accuracy to each marker and mesh smoothness at each local domain respectively. To seamlessly merge between an individual facial geometry and the occipital region's template mesh geometry, we set w_i^m and w_i^s as equation (3).

$$w_i^m = \begin{cases} 10\phi_i & if\ \phi_i > thr \\ 1 & else \end{cases} \qquad w_i^s = \begin{cases} 1 & if\ \phi_i > thr \\ 10(1-\phi_i) & else \end{cases} \tag{3}$$

Where, ϕ_i is equal to 1 if the scanned individual geometry is allocated to i-th vertex. In the other cases, ϕ_i is set to 0. The thr is the threshold for the weight ϕ. Based on our experiment, our system performs optimally when $thr = 0.8$. α and β decide contributions for fitting accuracy and smoothness effect. We minimize this energy function to estimate optimal affine matrices using L-BFGS-B, a quasi-Newtonian solver [17], and transform all of the vertices on the template mesh using the estimated affine matrices. As a result, we can obtain the complete head model (Fig. 1.(j)) that the both geometries are seamlessly merged (Fig. 1.(i)).

4 Facial Animation Synthesis

In this section, we describe how to synthesize facial animation which reflects personal characteristics using the individual head model generated using the technique described in Section. 3. In this paper, we mainly focus on a smile facial expression which frequently appears in daily life as a customization target to reflect personal characteristics to their CG characters. The detail of each process is described step by step in the following sections.

4.1 Capturing Video with Subject's Facial Expression

The smile facial expression is known to be expressed by contracting the the *risorius* and the *zygomaticus minor-major* facial muscles anatomically. We therefore need to adjust the *risorius* and the *zygomaticus minor-major* facial muscles for each individual. In Paul Eckman's Facial Action Coding System (FACS), "Lip Corner Puller" and "Cheek Raise" are related to the *risorius* and the *zygomaticus minor-major* respectively. Therefore, we capture an individual in facial expressions loosely based on "Lip Corner Puller" and "Cheek Raise" using a high definition video camera (Sony HDR-SR12) with 29.97 fps. For the subsequent processing, we made individual firstly express own neutral face, and then perform both facial expressions. Several frames in captured videos are shown in Fig. 2.

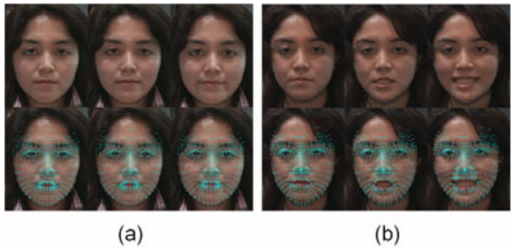


Fig. 2. The top row shows captured facial expression videos; the bottom row shows the tracking results. (a) and (b) are correspond to subject's "Lip Corner Puller" and "Cheek Raise".

4.2 Extracting Facial Motion Vectors

To adjust above three facial muscles for each individual, we need to extract person-specific motions from captured facial expression videos. On an initial frame of the captured video, the generic face mesh as shown in Fig. 3.(a) is fitted to the subject's face using the same manner in [9]. We refer to all of vertices of the resulting mesh as marker points. Then, maker points are tracked by the Lucas-Kanade Optical Flow Algorithm [5] overall the video (Fig. 2). After this tracking, we calculate facial motion vectors by subtracting marker positions of the initial frame with neutral from those of the end frame for each facial expression.

However, the magnitude and orientation of facial motion vectors are strongly depend on individual face geometry, so the motion vectors shouldn't be used to adjust facial muscles directly. To remove this dependency, we normalize individual face geometry to the generic face's one before facial motion vectors calculation. This normalization can be performed by mapping subject's marker positions with both neutral and facial expression to the generic face using Radial Basis Functions decided from the correspondence between marker points of subject's neutral face and those of the generic face. Finally, facial motion vectors using the normalized marker points are re-calculated.

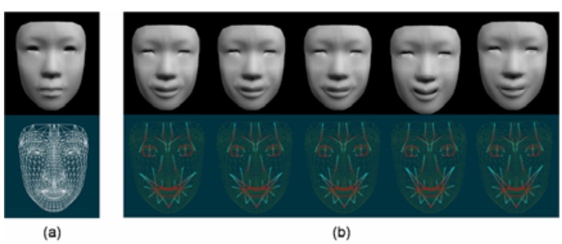


Fig. 3. (a)The generic face mesh (top: smooth shading, bottom: wireframe) (b) Example of simulated faces with different facial muscle layouts

4.3 Adjusting Facial Muscles for Each Individual

In this paper, we adjust a facial muscle layout for each individual to represent person-specific facial expression. The facial muscle adjustment process is composed by offline learning step and online estimation step. For learning step, using the generic face, 8 faces are generated by facial muscle simulation [15] with 8 different facial muscle layouts and common muscle contraction parameters (Fig. 3.(b)). Then, facial motion vectors for each simulated face are calculated by subtracting the vertex positions of the generic face with neutral expression from those of simulated faces. To focus on areas with large variance of facial motions, the galley of calculated facial motion vectors is applied to Principal Component Analysis (PCA). Each facial motion vector is mapped onto the PCA subspace spanned by only first eigenvector with the largest variance of facial motions.

For estimation step, first, an acquired subject's facial motion vector is mapped onto the PCA sub-space. Second, we calculate distances between the principal component of subject's facial motion vector and that of each simulated facial motion vector using Euclidian distance metric. Finally, the individual facial muscle layout is determined by finding the facial muscle layout with the minimum distance. These processes are separately performed for "Lip Corner Puller" and "Cheek Raise" cases. Consequently, we have 64 variations of facial muscle layouts for representing individualized facial expressions.

4.4 Synthesizing Facial Animation with Individualized Key Shapes

Using the estimated facial muscle layout, individualized key shapes are generated by physics-based simulation [15]. To accelerate this process, we adopt a multi-resolution approach, i.e, the physics-based simulation is performed with low-resolution face mesh (Fig. 3.(a)), and then the simulation results are retargeted to the individual head model by Facial Motion Distribution Chart technique [16]. The generated key shapes are shown in Fig. 4. As a result, person-specific facial animations are synthesized by blending individualized key-shapes according to pre-defined blend weight sequences in real-time [9].

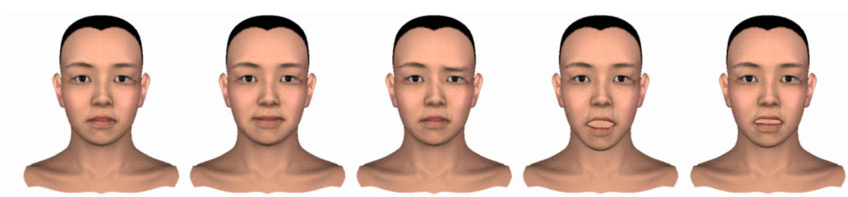


Fig. 4. Several examples of individualized key shapes

5 Result and Discussions

We implemented the proposed head modeling system using C++, GNU scientific library version 1.8, OpenCV library and L-BFGS-B library [24]. Then, we created

100 head models using the template mesh fitting framework proposed by Allen et al [2] (We will refer to this method as the Allen'03), and 100 using ours using an Intel Core 2 Duo T9600 2.8GHz.

Fig. 5 shows the performance comparison of the proposed and the Allen'03's fitting technique. The left of the second row represents the target surface which is polygonized by connecting edge between each marker. The middle one shows the face model generated using the Allen'03. Note that, the weight of the stiffness term in the energy function set to 10.0 during the final stage of the iteration so that artifact wouldn't occur.

The right one is the face model generated using our method. Our method succeeds to reduce distortions on the boundary between the scanned geometry and the template one as shown in Fig. 5. The third row is the RMS which describes the Root-Mean-Square distances between markers on the target surface and vertices on the generated face model. Additionally, the value in parenthesis represents the percentage of the RMS if the distance between eyes equal to one. As a result, fitting accuracy of both methods are almost same (no significant difference).

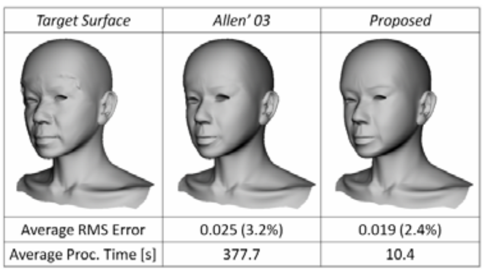


Fig. 5. The performance comparison between the previous and our techniques

The average computational times for each technique were 377.7 and 10.4 seconds respectively. Note that, in this experiment, we had no restriction of the number of the iteration for the optimization. Our template mesh fitting has advantage that the number of iteration required to the optimization can be reduced by fixing markers corresponding to the all of vertices of the template mesh. Therefore, our method is similar to the Allen'03 from fitting accuracy perspective, and is 36.3 times faster than it.

To demonstrate the effectiveness of our system, we implemented the proposed techniques as part of the Future Cast System in the National Museum of Emerging Science and Innovation in Tokyo. After presenting the Future Cast movie (like a Fig. 6) to visitors, we carried out the questionnaire survey "How clearly could you recognize yourself in the film?" for all of visitors. During this trial, a total of 147 subjects experienced the new version of the Future Cast System. As the result of the questionnaire survey, we found that 86% subjects could identify themselves. This score is 21% higher than our previous work [9]. This is because our new head modeling system can generate a more individualized CG character with person-specific head geometry and facial expressions.

Furthermore, our system can quickly generate an individual head model and key shapes required for facial animation within approximately 2 minutes (5 seconds for

Fig. 6. Snapshots from the Future Cast movie

scanning an individual facial geometry 30 seconds for generating an individual head model, 5 seconds for acquiring facial expression videos, 36 seconds for facial muscle estimation, 37 seconds for generating key shapes). This modeling time is comparable to the previous version of the Future Cast System [9]. Therefore, we conclude that the proposed techniques are effective to an entertainment system like a Future Cast.

Acknowledgements. We would like to thank to Hiroto Yarimizu, Takanori Suzuki and Yasushi Ishibashi for development of the facial animation framework. This work was supported by "R&D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Society", Special Coordination Fund for Promoting Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government, Grant-in-Aid for Scientific Research "High Fidelity Face Model for Analysis and Synthesis of Personal and Aging Characteristics", and Grant-in-Aid for Young Scientists (B) 21700207.

References

1. Alexander, O., Rogers, M., Lambeth, W., Chiang, N., Devebec, P.: Creating a Photoreal Digital Actor: The Digital Emily Project. In: SIGGRAPH 2009: ACM SIGGRAPH 2009, Course Notes (2009)
2. Allen, B., Curless, B., Popović, Z.: The Space of Human Body Shapes: Reconstruction and Parameterization from Range Scans. In: SIGGRAPH 2003: ACM SIGGRAPH 2003, Papers, pp. 587–594 (2003)
3. Choe, B., Lee, H., Ko, H.: Performance-driven Muscle-based Facial Animation. In: Proc. of Computer Animation, vol. 12, pp. 67–79 (2001)
4. Blanz, V., Basso, C., Poggio, T., Vetter, T.: Reanimating Faces in Images and Video. Computer Graphics Forum 22(3), 641–650 (2003)
5. Bouguet, J.-Y.: Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the Algorithm (2000)
6. De Carlo, D., Metaxas, D., Stone, M.: An Anthropometric Face Model using Variational Techniques. In: SIGGRAPH 1998: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, pp. 67–74 (1998)
7. Kähler, K., Haber, J., Yamauchi, H., Seidel, H.-P.: Head Shop: Generating Animated Head Models with Anatomical Structure. In: SCA 2002: Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 55–63 (2002)
8. Kraevoy, V., Sheffer, A.: Template Based Mesh Completion. In: SGP 2005: Proceedings of the third Eurographics Symposium on Geometry Processing, pp. 13–22 (2005)

9. Maejima, A., Wemler, S., Machida, T., Takebayashi, M., Morishima, S.: Instant Casting Movie Theater: The Future Cast System. IEICE Transactions on Information and Systems E91-D(4), 1135–1148 (2008)
10. Mukai, T., Kuriyama, S.: Geostatistical Motion Interpolation. In. In: SIGGRAPH 2005: ACM SIGGRAPH 2005, Papers, pp. 1062–1070 (2005)
11. Pighin, F., Hecker, J., Lischinski, D., Szeliski, R., Salesin, D.H.: Synthesizing Realistic Facial Expressions from Photographs. In: SIGGRAPH 1998: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, pp. 75–84 (1998)
12. Joshi, P., Tien, W.C., Desbrun, M., Pighin, F.: Learning Controls for Blend Shape based Realistic Facial Animation. In: SCA 2003: Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 187–192 (2003)
13. Sifakis, E., Neverov, I., Fedkiw, R.: Automatic Determination of Facial Muscle Activations from Sparse Motion Capture Marker Data. In: SIGGRAPH 2005: ACM SIGGRAPH 2005 Papers, pp. 417–425 (2005)
14. Vlasic, D., Brand, M., Pfister, H., Popović, J.: Face Transfer with Multilinear Models 24, 426–433 (2005)
15. Lee, Y., Terzopulus, D., Waters, K.: Realistic Modeling for Facial Animation. In: Proceedings of SIGGRAPH 1995, pp. 55–62 (1995)
16. Yotsukura, T., Morishima, S., Nakamura, S.: Face Expression Synthesis based on a Facial Motion Distribution Chart. In: SIGGRAPH 2004: ACM SIGGRAPH 2004, Posters, p. 85 (2004)
17. Zhu, C., Byrd, R.H., Lu, P., Nocedal, J.: Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. ACM Trans. Math. Softw. 23(4), 550–560 (1997)

Geppetto: An Environment for the Efficient Control and Transmission of Digital Puppetry

Daniel P. Mapes, Peter Tonner, and Charles E. Hughes

University of Central Florida, 4000 Central Florida Blvd.
Orlando, FL 32816-2362 USA
dmapes@ist.ucf.edu, ptonner@knights.ucf.edu, ceh@cs.ucf.edu

Abstract. An evolution of remote control puppetry systems is presented. These systems have been designed to provide high quality trainer to trainee communication in game scenarios containing multiple digital puppets with interaction occurring over long haul networks. The design requirements were to support dynamic switching of control between multiple puppets; suspension of disbelief when communicating through puppets; sensitivity to network bandwidth requirements; and as an affordable tool for professional interactive trainers (Interactors). The resulting system uses a novel pose blending solution guided by a scaled down desktop range motion capture controller as well as traditional button devices running on an standard game computer. This work incorporates aspects of motion capture, digital puppet design and rigging, game engines, networking, interactive performance, control devices and training.

Keywords: Digital puppetry, avatar, gesture, motion capture.

1 Introduction

Digital puppetry refers to the interactive control of virtual characters [1]. The artistry of using a human to control the actions of animated characters in real-time has been employed for several decades in a number of venues including television (Ratz, Muppets [2]) and location-based entertainment (Turtle Talk with Crush [3]). More recently, these techniques have found application in interactive networked simulations [4,5].

By affording responsive and adaptive behaviors while operating through the changeable context of an invented digital character, puppetry greatly increases the effectiveness of interactive experiences. Not unexpectedly, these benefits come with demanding requirements both from the puppeteers and from the technical aspects of projecting the simulation. From the puppeteer, there is the cognitive load associated with rapid switches between puppets; maintaining individual puppet behaviors during switches; following training scenario scripts; operating scenario specific controls; and operating puppet specific behavior controls. From a technical aspect, there is the need to reliably and efficiently capture and transmit behaviors over a long haul network; the necessity to achieve this transmission even in the presence of limited resources,

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 270–278, 2011.

including bandwidth and computational power; and the requirement to articulate these actions at the client site, providing smooth interactive performance.

Fig. 1. Pashtun avatars used in cross-cultural training

In our paper we present an evolution of interaction paradigms that have made projected puppetry more intuitive and less cognitively draining while reducing network bandwidth requirements through a micro pose blending technique guided by body worn 3D controllers. The use of pose blending reduces visual artifacts of live motion captured puppetry; supports culturally appropriate gestures; and significantly reduces the streaming transmission bandwidth requirements over long-haul networks. Essentially, our contribution is interactive, affordable, reliable, believable, culturally attuned long-distance digital puppetry.

2 Social Contexts and Puppetry

2.1 FPS Gaming

One of the most prevalent and affordable examples of projected puppetry applications is the networked first person shooter game where individual players log in online to control a puppet that can do a fixed set of behaviors such as walk, run, duck, jump, punch, kick, aim and shoot using devices such as game controllers, keyboards and mice. While the process of pressing a button to control a character may not be initially intuitive, devoted players to these games have become quite adept at using these controls under the pressure of competitive game play. The control data being transmitted over long-haul worldwide networks from the buttons and dual axis joysticks found in typical controllers is insignificant compared to the incoming game state stream or the streaming VOIP that players use to collaborate during play.

The ability for players to collaborate in teams led to the emergence of Machinima where remote participants could interact with each other to create short playful movies where each player acted out a predefined role, while the results were recorded through the game camera. Though the control model allowed for small efficient network streaming, with only the ability to run, walk, duck, jump, aim and shoot it was fairly clear that the game characters lacked the more complex facial expressions and body language to do live acting.

2.2 Second Life

In an online social networking game like Second Life over 120 user-playable puppet animations are provided by default with the option of players increasing this set with their own custom animations [6]. These are used much like emoticon symbols embedded in a text document. While this is an improvement over the FPS game control by showing both facial emotion and body language, even if a puppeteer were able to master such a sizable abstract mapping, the performance still lacks the subtle hand and body motions used to communicate non-verbally and to punctuate speech.

3 Evolution of Digital Puppetry in Geppetto

3.1 Pre-service Teacher Training

The TeachME™ experience provides pre-service and in-service teachers the opportunity to learn teaching skills in an urban middle school classroom composed of five puppets, each with its own prototypical behavior. Teachers use the virtual classroom to practice without placing “real” students at risk during the learning process [4]. Overall, the concepts developed and tested with the TeachME environment are based on the hypothesis that performance assessment and improvement are most effective in contextually meaningful settings.

In the design of TeachME the puppets (avatars) were always seated at desks which allowed us to focus only on upper body motions. We captured head, shoulder, elbow and waist orientations by placing 9 retro-reflective markers on the puppeteer and tracking them as 3D points using a 4 camera IR motion capture system. Four basic facial poses, smile, frown, wink and mouth open were simultaneously blended using the finger bend sensors of a CyberGlove [7]. Lip sync through audio evaluation was tried and rejected due to excessive evaluation period latency and feedback from the puppeteers on their need for more direct control. An Ergodex game pad was used to allow the puppeteer to rapidly switch control into one of the five middle school age puppets and provide a custom key interface to control the scenario. A Skype connection was opened up to provide bi-directional audio between the trainer and trainee as well as a trainer webcam video feed of the trainee.

3.2 Avatar

The Avatar project used digital puppets (Figure 1) to practice cross cultural communication skills. Two scenarios were created, one in the Pashtun region of Afghanistan and one in an urban setting of New York. Each scenario had three puppets: one elder leader, a young idealist, and an antagonist. In each scenario the participant was to practice successful negotiation with the trained Interactor operating through the puppets.

In order to minimize the limitations from live mocap in the TeachME project, a micro-pose system was designed where each puppet contained a set of full body end poses (Figure 2) that were characteristic of how each puppet might communicate through hand and body posture. A scaled down dual IR camera motion capture system was used to track the Interactor's head orientation and the location of the actor's wrists. A calibration process was created where each characteristic puppet pose was presented to the Interactor who was then asked to mimic the pose. A database of their real world wrist locations and the puppet's wrist locations was built and a routine was written to use the Interactor's wrist locations to find the two closest poses from the puppet with a confidence weight on each pose. These two poses were then blended in proportion to their weights along with previously used poses decaying in weight to zero over time. Facial pose was controlled using the same five-finger glove metaphor from TeachME. The Ergodex game controller was replaced with a WiiMote held in the non-gloved hand.

Fig. 2. Key poses for avatars

The end result was a puppet designed to exhibit only culturally appropriate gestures where these gestures could be controlled seamlessly using 3D locators held in each of the Interactor's hands while the Interactor moved the puppet's head directly by their own head movement. Since the puppet poses and the controlling hand locations were mapped nearly one-to-one, this system was almost as intuitive as raw mocap but with fewer motion artifacts.

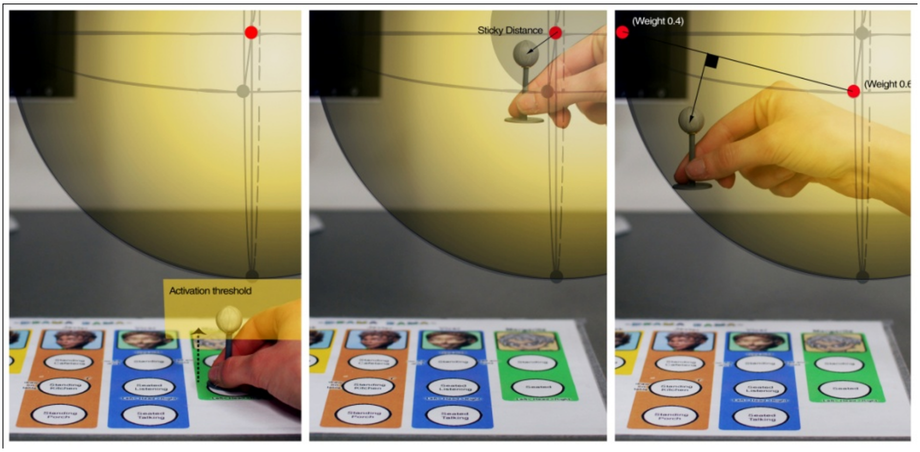
The Avatar effort improved the speed of puppeteering by feeding a much smaller vector of data points (two wrist locations) into the animation database. This addressed issues in previous studies where motion capture was stored in a database of animations and full motion capture was used to map into this database in real-time. Incorporating more animations would require scaling the database appropriately and optimizations to mapping of the mocap data to these animations [8]. In addition, all aspects of the pose, including hand and finger postures, could be included where the raw mocap system had only flat unexpressive hands. Another clear benefit of this system over raw motion capture was the realization that only the poses and their relative weight need be communicated over the Internet to describe the puppet state. Instead of streaming 17.28 mbps of skeletal joint angle data, this pose blended system required on the order of 1.44 mbps. While this is not significant for projecting a single puppets performance, when social networking games and machinima servers emerge supporting 100's of puppets the need for performance compression will become more apparent.

After the initial trials with this system, with only 3D points at their wrists, the puppeteers complained that the selection criteria was too ambiguous to find some of the key gestures under the pressure of performance. An example might be a hand held palm up or palm down. Another minor problem was the need to maintain individual calibration settings for each puppeteer as they changed shifts.

3.3 Evolving Avatar

In an effort to further minimize the need for full body motion capture and give puppeteers the control they requested, the representative poses from the initial study were sorted into bins based on modes within which the puppeteers were operating. Many of the poses were designed to be used together and, once separated, the blending became trivial to control with two points; but now a method for choosing which bin to operate within needed to be found. Button events from either an Ergodex or WiiMote were considered but this added to the buttons already being used to control the scenario and to switch between puppets.

The solution that evolved was a hybrid between motion capture and a game controller solution, where the need to switch between puppets and different puppet pose spaces were combined. A single 3D point from the motion capture system was used as a controller within a virtual 3D volume above a desk surface. A printed map of each of the pose bins of all of the puppets was attached to the desk surface. Placing the marker below a fixed height above the desk surface put all puppets into an automated non-puppeteered state (Figure 3a) and lifting the marker above a pad designated which puppet was to be controlled as well as the set of poses to be blended. When lifted above a pad the controller values were converted to a vector relative to a sphere of fixed radius and centered above the pad. The vector was



(a) (b) (c)

Fig. 3. Using marker to select character and behavior genre and then effect behaviors

normalized to have a length of 1.0 at the sphere surface. The poses within each bin were then assigned a representative vector within the sphere coordinates that best expressed the nature of the pose relative to other poses in the bin. A vector pointing up (0,1,0) might be mapped to a pose with arms above the head or a character making the thumbs up gesture. Sometimes poses progressed along an axis. For instance a left moving gesture might have a critical pose representation at (1,0,0) followed by another at (2, 0, 0). A design workflow quickly developed between the puppeteers and the artists making poses where they expressed their requests and results as up, down, left, right, forward and back with extended poses as up-up, down-down etc. While there was no real restriction on the use of vectors like (0.5, 1.0, 0.0) , the human element in the design process did not work well with abstract numbers either in design or control. Another nice aspect of this design workflow was that it allowed artists to work from representative pictures provided by cultural content experts involved in the project. Once a puppet was rigged for animation it was an easy matter to create and edit poses based on expert feedback.

Finding the two closest poses within a bin was done by picking the two closest pose vectors to the controller, a process which no longer required individual user calibration. The active two poses were weighted proportional to each other when transitioning from one to the other by projecting the 3d marker to the line between each pose’s control point. When the projected controller fell halfway between each pose, the weights were 0.5 and 0.5 respectively (Figure 3c). When the projection fell past either of the endpoints the closest endpoint was weighted at 1.0 and the other at 0.0. When a previously active pose became deselected its current weight was decayed to 0.0 at a fixed rate configured by the puppeteers. In some cases when a pose was spatially organized in between other poses a radius was added to the endpoint which gave a “sticky” zone where the pose was held (Figure 3b).

This new system gave a great deal of control over the appropriate cultural feel of the puppet, gave the puppeteer much better control of both the puppets they would control and also the set of poses they wished to use. The pose blending now was operated by a single hand with the elbow comfortably resting either on the desk or chair armrest, leaving the other hand to control the scenario and facial poses. The head orientation could still be directly controlled by the same IR camera system measuring the control marker. Compared to previous studies where puppet actions would be visibly displayed, and then require costly comprehension by the user [Lee et al., 2002], an Interactor in this system can store possible actions mentally and draw from these possibilities in real time.

3.4 DramaRama

This study is developing and testing an interactive virtual environment that is designed to build the skills in middle school girls needed to resist peer pressure. Figure 3 depicts a prototype of the control pad used by puppeteers in this study. Along the top are the five characters that appear in the set of scenarios. Below each character is a set of behavior types (Figure 4).

This project was the first to use the enhancements to the Geppetto system derived from the Avatar project trials. Anywhere from 1 to 3 bins were created for each of the 5 puppets. The bins for each puppet were arranged left to right according to where each puppet typically appeared on screen relative to the other puppets. Multiple bins for a puppet were stacked vertically top to bottom based with standing pose sets on the top followed by seated variations. Printed images of each puppet titled each column and descriptive annotations for each bin provided additional feedback during initial training.

Head pose from the full body blend was further modulated by the orientation of the head measured using an IR marker pattern mounted on a cap worn by the puppeteer. A Logitech G13 advanced gameboard was used to control facial pose blending of three expression-pair opposites and the mouth open pose. As the key for any pose was held the weight of the gesture would grow until it reached full weight at a rate specified by the puppeteers. When released the pose would lose influence at a weight decaying slower than initially used for rising to allow rich blending. The mouth open pose had a faster rise and decay time to allow for better lip sync. The quality of the key approach over the glove was in part to simplify the setup. In comparison, the glove required calibration and was generally less reliable. A novel lip tracker using 4 IR markers attached around the puppeteers mouth was explored but the puppeteers preferred the simplicity and control afforded by the gameboard. What was important in all approaches was the ability to create weighted blends involving all facial poses simultaneously. Having multiple poses mixing over time allows Geppetto to transcend the plastic look often associated with digital puppets, thereby bringing the character to life. While this was not done, a certain amount of random background noise is suggested from this result. There was an automated blink pose on all puppets that also added a great deal to the overall effect.

The result is a lifelike, almost infinitely varied set of gestures that, together with interactive dialogue, that provides deep immersion on the part of the user.

Fig. 4. Characters in DramaRama

4 Conclusions

The primary gains of this approach to puppetry over full-body capture have been the reduction in noise that arises from tracking a large number of points; a reduced hardware footprint; reduction in overall hardware cost and a significant reduction in data that must be transmitted; Additionally, the approach greatly reduces the cognitive and physical load on our puppeteers, especially when they must be concerned about varying cultures to which this must be delivered – non-verbal cultural awareness exists at the receiving end not at the puppetry end.

Acknowledgements. The material presented in publication is based on work supported by the National Institute for Nursing Research (R15NR012189-01) and the National Science Foundation (CNS1051067). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Institute for Nursing Research or the National Science Foundation.

References

1. Sturman, D.J.: Computer Puppetry. Computer Graphics and Applications 18(1), 38–45 (1998)
2. Walters, G.: The Story of Waldo C. Graphic. Course Notes: 3D Character Animation by Computer. In: Proceedings of the 16th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1989), Boston, pp. 65–79. ACM Press, New York (1988)
3. Trowbridge, S., Stapleton, C.: Melting the Boundaries between Fantasy and Reality. Computer 42(7), 57–62 (2009)
4. Dieker, L., Hynes, M., Hughes, C.E., Smith, E.: Implications of Mixed Reality and Simulation Technologies on Special Education and Teacher Preparation. Focus on Exceptional Children 40, 1–20 (2008)
5. Mazalek, A., Chandrasekharan, S., Nitsche, M., Welsh, T., Thomas, G., Sanka, T., Clifton, P.: In: Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games, New Orleans, Louisiana, pp. 161–168. ACM Press, New York (2009)
6. Rymaszewski, M.: Second Life: The Official Guide. Wiley-Interscience, New York (2007)
7. Kessler, G., Hodges, L., Walker, N.: Evaluation of the CyberGlove as a Whole-Hand Input Device. ACM Transactions on Computer-Human Interactaction 2(4), 263–283 (1995)
8. Lee, J., Chai, J., Reitsma, P.S.A., Hodgins, J.K., Pollard, N.S.: Interactive Control of Avatars Animated with Human Motion Data. In: Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2002), San Antonio, pp. 491–500. ACM Press, New York (2002)

Body Buddies: Social Signaling through Puppeteering

Magy Seif El-Nasr[1], Katherine Isbister[2], Jeffery Ventrella, Bardia Aghabeigi[1], Chelsea Hash, Mona Erfani[1], Jacquelyn Morie[3], and Leslie Bishko[4]

[1] Simon Fraser University
[2] New York University-Poly
[3] Emily Carr University
[4] University of Southern California

magy@sfu.ca, isbister@poly.edu, Jeffrey@ventrella.com, b.aghabeigi@gmail.com, saveremreve@gmail.com, morie@ict.usc.edu, lbishko@ecuad.ca

Abstract. While virtual worlds have evolved to provide a good medium for social communication, they are very primitive in their social and affective communication design. The social communication methods within these worlds have progressed from early text-based social worlds, e.g. MUDS (multi-user dungeons) to 3D graphical interfaces with avatar control, such as Second Life. Current communication methods include triggering gestures by typed commands, and/or selecting a gesture by name through the user interface. There are no agreed-upon standards for organizing such gestures or interfaces. In this paper, we address this problem by discussing a Unity-based avatar pupeteering prototype we developed called Body Buddies. Body Buddies sits on top of the communication program Skype, and provides additional modalities for social signaling through avatar pupeteering. Additionally, we discuss results from an exploratory study we conducted to investigate how people use the interface. We also outline steps to continuously develop and evolve Body Buddies.

Keywords: avatar pupeteering, avatar nonverbal communication, social communication with avatars, avatar design, CVE (Collaborative Virtual Environment).

1 Introduction

Mobile devices, including phones and PDAs, are becoming the dominant method for communication and an essential part of our everyday lives. Communication and social interaction have shifted from standard face to face modality to mediated social settings, such as email, Facebook, Skype videoconferencing, and increasingly, online multi-user virtual worlds. Despite the increased technical enhancements making such synchronous and asynchronous communication modalities possible, the design of synchronous online communication systems is still limited and primitive in terms of the affordances for social interaction and the affective communication they offer.

Extensive research in areas, such as communication and social psychology, has highlighted the significance of nonverbal behaviors, such as facial expressions, turn

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 279–288, 2011.

taking signals, and body language for communication [1-3], [4]. Current synchronous communication systems use one or more of four communication modes: text, audio, video, and avatar systems. Chat or text-based interfaces are limited since users cannot communicate intricate social messages, such as turn taking, or signals of skepticism or confusion. Recently, there are systems proposed or developed to combine video and audio signals, such as Skype calls, or PortaPerson [5]. Such modalities enable users to communicate and deliver synchronous social and affective messages through non-verbal behaviors within audio and video channels. These are successful solutions for one-on-one settings; however, there are issues that constrain the use of such systems within a group mode. First, with video alone it is impossible to use gaze direction as a communicative element between more than two people. Even with a setup as the one discussed in [5], it is still hard to efficiently use gaze; though it is widely recognized as an important communicative element [6], [7]. Second, spatial placement is hard to communicate through video, especially because people are not co-located in their communication space. A better method for enabling body position and proximity is to use virtual environments or spaces with an avatar representing the user so each person in a group is co-located virtually. Several researchers explored capturing gestures and developing avatars that imitate user gestures in virtual space [8], [9]. This approach has several limitations including methods of using or transferring proximity and spatial signals. In this paper, we take a different approach.

In the past year, we formed an interdisciplinary team composed of designers, developers, artists, a graphic designer, and a communication researcher to address this issue. We developed a Unity-based avatar puppeteering system, called *Body Buddies*, that sits on top of the Skype conferencing application, allowing Skype users to socially signal messages to one another in a simple virtual environment. Avatars can be adjusted to show like/dislike, skepticism, agreement, attention, and confusion, using dynamic movement rather than static poses. The system was first demonstrated and published at the CHI 2010 workshop on social connectedness [10]. In developing *Body Buddies*, we focused on allowing users conscious control of various puppeteering parameters. This approach has various advantages and disadvantages. First, it requires the user to consciously make a signal. Second, it adds cognitive load on the user as they take the burden to communicate these signals when needed. However, using this type of interface allows users more control over their signaled behaviors. It also may alleviate video camera issues, such as users not wanting their image to be projected or feeling nervous in front of a camera [11].

In this paper, we discuss the system and its puppeteering interface. In addition, we discuss preliminary results of a study we conducted, in which we asked users to discuss and debate a particular topic using *Skype* and *Body Buddies*. We conclude the paper by discussing future research.

2 Previous Work

Previous work within this area spans multiple disciplines. We outline the following areas: *nonverbal behavior in real life*, for which we devote a section discussing models and taxonomies proposed. It should be noted that in the interest of space, we only summarize some important contributions here, highlighting what models transfer to Virtual Worlds (VWs). The second relevant area is: *avatar based online communication environments*. Although there has been little work in this area, we highlight some of the significant work here. Some of this work includes systems and/or studies of how people used avatar based nonverbal communication modalities within a virtual environment. These studies are an important corner stone to our work.

2.1 Nonverbal Behavior

The study of nonverbal behavior in the real world has received much attention, including the study of proximity, emotional expressions, and gesture. Hall and Birdwhistell [1], [12], [13] are considered the fathers of the study of *Proxemics* and *Kinesics,* respectively – two of the very important and dominant paradigms of nonverbal communication dealing with different aspects of the human body.

Hall's work on *Proxemics* discusses the notion of personal space, describing several zones of intimacy around the body. Over its 60 year history, Proxemics has been used to describe how people position themselves in space relative to each-other, and how different demographic factors alter these spatial behaviors. Recent studies in Virtual Worlds (VWs) discussed evidence found supporting the translation of real world proxemic and gaze behavior to virtual worlds [6], [14], [15]. Additionally, Yee et al. also report the presence of social norms governing the proxemic behaviors within virtual worlds resembling those of the real world [16].

Kinesics, which is the study of gesture and posture, has also received attention. In addition to the structural model developed by Birdwhistell [13], several researchers investigated a descriptive approach. Ekman and Friesen [2] present an exhaustive description of the types of non-verbal behavior that people perform. They discuss different types of acts, such as *emblems*: culture specific, learned behaviors that represent meaning, *illustrators*: socially learned behaviors that complement or contrast verbal messages, *affect displays*, *regulators*: conversational flow gestures that control the back and forth within a dyad, and *adaptors*: learned actions based on satisfying bodily needs, based on child-hood experience. This model has been used by several researchers within the HCI field [8].

Additionally, there has been much work on the use of gesture in speech and communication. An important work in this area is the work of McNeill and Cassell [4], [17], [18], who explored the use of communicative gestures by observing and analyzing people talking about specific subjects, such as real estate, etc.

2.2 Avatar-Based Nonverbal Communication within Online Meeting Environments

Several researchers empirically investigated the communicative power of nonverbal behaviors within virtual environments. In a study conducted by Allmendinger, they compared conditions with video, audio, inferred-gaze avatar, and random-gaze avatar. They found that video was most favored followed by inferred-gaze avatar system [11]. This confirms the role of gaze in nonverbal communication as discussed in previous work [6], [7]. Additionally, automated gaze within avatar groups were explored and implemented in the socially-focused virtual world There.com. Through in-house user testing, such use of gaze was found to significantly increase users' sense of social engagement [3]. In addition to gaze, Allmendinger argued that avatars can provide cues to support (a) *group awareness*, such as focus attention and position in an argument, as well as (b) *communicational gestures*, such as signals to identify who is talking [11].

Empirical work exploring the design of such avatars is sparse, although there are some. Anderson et al. presented a study combining testing and participatory design to infer the usability and presence of avatar systems. Their experiments showed that users needed a level of control on avatar animation to show who is talking and support control and turn-taking [19]. Similarly, Guye-Vuilleme et al. [20] stressed the use of agreement and space in avatar design; they deduced these results through a qualitative experiment with a CVE (Collaborative Virtual Environments). This is important as turn-taking in distributed synchronous environments is seen as a problem area [21], [22]. Allmendinger et al.'s study confirmed these results by concluding that important signals for avatars in CVEs were: *thumb up*, *gestures highlighting information on slides*, and *turn taking signals* [23].

The work on developing virtual meeting spaces can be grouped into two groups: sensor-based intelligent environments, where gestures are entered through devices, camera or other sensors and are transferred to an avatar model [24], [25], and lightweight interactive virtual environments, such Lucia et al.'s *SLMeeting* [26] which supports collaborative and management activities within a web based interface, and *Porta-Person* [5], which enhances the sense of social presence through a controlled display of a video image of the participant. Another example of lightweight interactive virtual meeting system is Shami et al.'s Olympus [27], which is a flash-based light-weight virtual meeting place that allows chat based input, and can link specific commands to gestures. For example, the character can shrug based on the text '?'. Similar, to WOW interface, users can type specific gestures by typing '/' then the name of the gesture animation. They tested this system in three meetings to assess its effectiveness. In terms of the use of gesture, they found that in meetings users used a combination of gesture and chat in general. The most popular three gestures, confirming previous research, were: clap, agree, and wave. They, however, concluded that users did not move their avatars. Our work extends the work discussed here to present a new avatar puppeteering system and test its interface.

3 Body Buddies

The architecture of the Body Buddies system is shown in figure 1. The system consists of two components. The avatars along with the virtual environment were developed in Unity. The Unity-based system is developed to augment Skype voice interaction with avatar based controls. It is activated once the user logs into the Skype program. The Skype – Unity interaction is implemented using a middleware protocol, which acts as a mediator between the Unity client avatar system and the Skype program, transferring messages from Skype to Unity and vice versa. As shown in the diagram, the middleware application communicates with Unity through TCP/IP protocol and with the

Skype client through an ActiveX component representing Skype API as objects called Skype4Com. The middleware first establishes a connection between Unity and Skype through a handshaking routine. This is done for all users within a Skype call. The middleware application then uses the application-to-application protocol command in the Skype API, *AP2AP*, to send Unity messages from one Skype client to other peers in the conference call. For example, when a user activates a gesture command for an avatar in a client, the Unity side sends a message to the middleware using TCP/IP protocol, and the mediator uses the *AP2AP* commands for sending this animation change to other peers in the Skype conversation. On the other side of the communication, the Skype client communicates with the middleware side using a Skype4Com registered call back, and then the application sends the animation command to the other Unity part by TCP messaging system. Finally, the Unity side parses the animation command, and executes the related avatar changes.

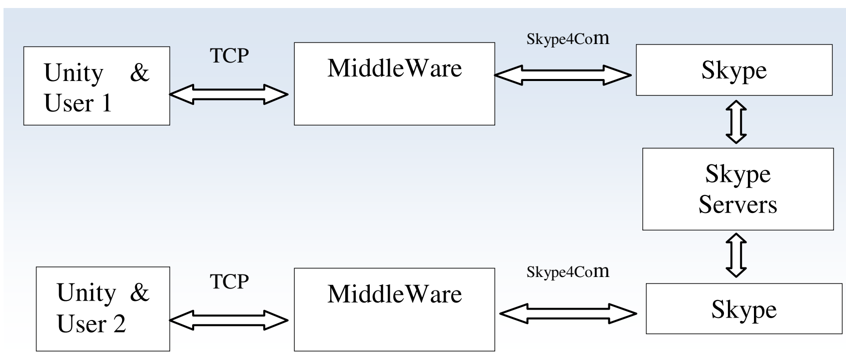


Fig. 1. Body Buddies Architecture

Once the user logs in, the Unity-based interface shown in figure 2 appears showing the user the other avatars with the Skype name for each avatar displayed above its head. The user can interface with the avatar through the buttons shown in figure 2, where he/she can move, rotate, lean the avatar forward or backward, as well as execute the social gestures 'skeptical' and 'my turn'.

The avatars in the *Body Buddies* system use a hybrid set of techniques. The 3D representations were modeled and rigged in Maya, and then imported into the Unity game engine, along with accompanying short-lived full-body gestural and postural animations. The UI controls for moving the avatar (forward, backward left and right) as well as *skeptical* and *My Turn!*, shown in figure 2, were linked to the Maya animations developed for the avatars, thus allowing users to trigger animations in real-time. In addition to these triggered animations, controls were implemented allowing the user's avatar root position and heading to be adjusted – permitting a rudimentary form of navigation. This allowed users to shift the positions of the avatars in relation to each other and also to face towards or away from each other – for the purpose of social signaling.

Fig. 2. Body Buddies interface developed in Unity. For a full video of a demo see: http://www.sfu.ca/~baa17/SkypeBuddy/DemoLeslie.avi

In addition, we also added modifiers to the avatar joint rotations allowing the user to adjust parameters such as *Arch forward* and *Arch backward* [28]. These were procedurally-generated postural stances involving several joints. These procedural modifications were layered on top of the avatar joint array such that they could be smoothly blended with any imported animation playing simultaneously. The blending of postural and gestural movement created a palette of body signals that the user could combine in a variety of ways.

4 Study

To investigate how users interacted with the avatar system, we ran a study with 9 groups of 2-3 participants. Unfortunately, due to technical difficulty and problems with videos we had to disregard data from 3 groups; thus, we analyzed only 6 groups for a total of 11 participants.

4.1 Procedure

Participants were invited for a debate session in the lab in pairs. Once they arrived, they were asked to sign a consent form and then asked to complete a survey designed to measure their social connectedness. We then took each participant to a different room equipped with a laptop or desktop computer running *Skype* and the *Body Buddies* system. We then asked each participant to discuss a given topic (social network and Facebook) using the Skype and Body Buddies interface. We did not enforce any specific interface use during the session, leaving them to chat freely using the given tools. We video taped their interaction session for later analysis. We also logged all their actions, including button presses, the amount of time the Unity window was active, their button-pushing frequency, etc. This interaction session lasted between 15-18 minutes. After the session, we asked participants to fill out a questionnaire and the social connectedness survey again.

4.2 Results

Figure 3 shows the total time vs. the time spent using the avatar interface. Our results show that the users employed the avatar interface considerably more than any of the other interfaces. This result is statistically significant.

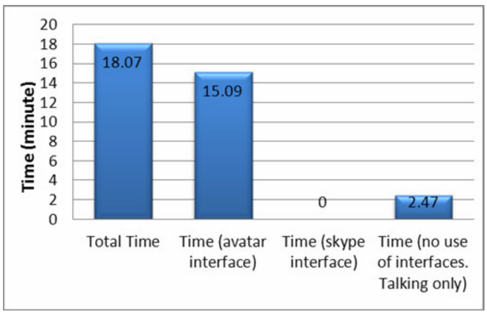
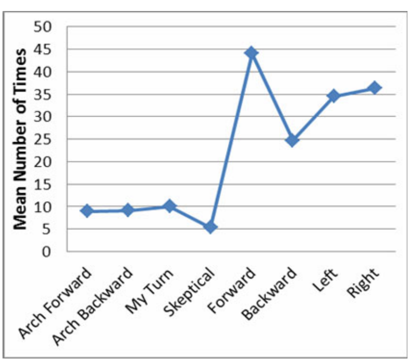


Fig. 3. Average time (error bars and bar chart)

Results from the before and after social connectedness tests show that IOS Scale upgraded on average from mean IOS Scale (before) = 3.53 to mean IOS Scale (after) = 4.4. While on average there is a difference, it was not significant — an expected result as 15 minutes is too short to cause a major improvement on social connectedness. However, interacting with avatars may improve social connectedness in the long run.



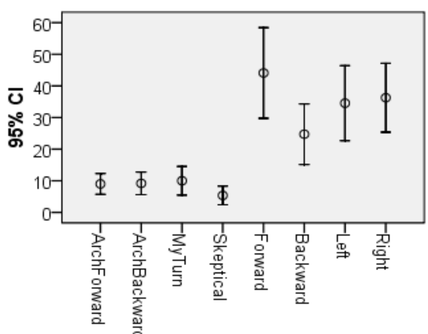


Fig. 4. Button press analysis. The figure on the right shows averages and the figure on the right shows error bars calculated from Standard Error given the sample.

The investigation of the interface use led to interesting results. The interface has 8 different buttons (Forward, Backward, Left, Right, Arch Forward, Arch Backward, My Turn and Skeptical). Analysis of the times these buttons were pushed, shown in figure 5, showed that participants used mostly the movement keys (Forward, Backward, Left and Right). It should be noted that the movement buttons were counted differently, as we only counted the first button press of a sequence of button presses. This is because users will probably press several times to move the avatar, to a specific place. We counted this as one event so as not to skew the data. We divide the eight buttons into two groups: movement buttons (Forward, Backward, Left and Right) and other buttons (Arch Forward, Arch Backward, My Turn and Skeptical). We ran Mann-Whitney U Test to determine any significance between the use of these two groups of buttons. The results show significance ($p<.05$). Participants pushed the movement buttons four times as much as the other buttons, which is an interesting result as it is in conflict with the other results from the literature discussed above, where Shami et al. [27] deduced that participants did not move avatars at all within meetings. Figure 4 shows the error bars on the average number of times of button presses and the average number of times of button presses over different actions.

In addition to this quantitative analysis, we also looked at the qualitative feedback given by users. First, some participants were enthusiastic with the addition of another layer of expressiveness, as one said, “being able to visually interact with other Skype users through emotions is great, when you do not wish to use your camera or do not have access to one.” But some expressed concerns, such as “it is difficult to concentrate on both moving the avatar around and talking at the same time.”

5 Conclusion and Future Work

The goal of this project was to investigate types of interfaces that could support better communication within computer-mediated meetings. We found, similar to previous work, that some affordances for avatar puppeteering were used more than others. Unlike previous work, we found that users used movement the most. However, the study is limited in several ways. The study was conducted in a lab setting with undergraduates. We think this limits the result as the behavior of participants in a real meeting versus a made up scenario will be different. Overall we saw participants were more interested in ‘playing’ with the system than communicating, perhaps due to its novelty or due to the setting itself. Therefore, we believe as we move on to a different setting for testing, the use of real meeting environments will be necessary to understand and investigate the use of nonverbal communication mediated by avatars.

Understanding the communicative affordances for the system is an interesting and complex problem. We suggest several future directions towards achieving this goal, including adding other social signals such as expressive emotions: ‘happiness’, ‘sadness’, etc., expressive confusion, greetings, thumb up and thumb down which previous literature has noted as important. We also hope to engage in additional investigations with the use of body animations, gestures, and postures as techniques for expressing these variables, and use of other devices beyond keyboard and mouse.

Acknowledgements

We would like to thank GRAND (Graphics Animation and New Media) – NCE (Network of Centers of Excellence) for making this work possible.

References

[1] Hall, E.: Proxemics. Current Anthropology 9, 83 (1968)
[2] Ekman, P., Friesen, W.: The repertoire of nonverbal behavior: Categories, origins, usage, and coding. Semiotica 1, 49–98 (1969)
[3] Ventrella, J.: Virtual Body Language. Eyebrain Books (2011)
[4] McNeill, D.: Gesture and thought. University of Chicago Press, Chicago (2005)
[5] Yankelovich, N., Simpson, N., Kaplan, J., Provino, J.: Porta-person. In: CHI 2007 extended abstracts on Human factors in computing systems - CHI 2007, San Jose, CA, USA, p. 2789 (2007)
[6] Bailenson, J.N., Beall, A.C., Blascovich, J.: Gaze and task performance in shared virtual environments. The Journal of Visualization and Computer Animation 13, 313–320 (2002)
[7] Colburn, A.R., Cohen, M., Drucker, S.M.: The Role of Eye Gaze in Avatar Mediated Conversational Interfaces, Microsoft Research (2000)
[8] Vinciarelli, A., Pantic, M., Bourlard, H., Pentland, A.: Social signal processing: state-of-the-art and future perspectives of an emerging domain. In: Proceeding of the 16th ACM international conference on Multimedia, pp. 1061–1070. ACM, New York (2008)
[9] Vinciarelli, A., Pantic, M., Bourlard, H.: Social signal processing: Survey of an emerging domain. Image and Vision Computing 27, 1743–1759 (2009)
[10] Isbister, K., Seif El-Nasr, M., Ventrella, J.: Avatars with Improved Social Signaling. In: CHI 2010 Workshop on Designing and Evaluating Affective Aspects of Sociable Media to Support Social Connectedness (2010)
[11] Allmendinger, K.: Social Presence in Synchronous Virtual Learning Situations: The Role of Nonverbal Signals Displayed by Avatars. Education Pyshcology Review 22, 41–56 (2010)
[12] Hall, E.T.: The Hidden Dimension. Anchor (1990)
[13] Birdwhistell, R.L.: Introduction to kinesics: An annotation system for analysis of body motion and gesture. University of Michigan Library (1952)
[14] Bailenson, J., Beall, A., Blascovich, J., Raimundo, M., Weisbuch, M.: Intelligent Agents Who Wear Your Face: Users' Reactions to the Virtual Self. In: Intelligent Virtual Agents: Proceedings of Third International Workshop, IVA 2001, Madrid, Spain, September 10-11, p. 86 (2001)
[15] Friedman, D., Steed, A., Slater, M.: Spatial Social Behavior in Second Life. In: Intelligent Virtual Agents, pp. 252–263. Springer, Paris (2007)
[16] Yee, N., Bailenson, J., Urbanek, M., Chang, F., Merget, D.: The Unbearable Likeness of Being Digital: The Persistence of Nonverbal Social Norms in Online Virtual Environments. Cyber. Psychology & Behavior 10, 115–121 (2007)
[17] Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., Stone, M.: Animated conversation: rule-based generation of facial expression, gesture \& spoken intonation for multiple conversational agents. Computer Graphics 28, 413–420 (1994)
[18] Cassell, J.: A Framework For Gesture Generation And Interpretation. Computer Vision in Human-Machine Interaction, 191–215

[19] Anderson, J., Ashraf, N., Douther, C., Jack, M.A.: Presence and Usability in Shared Space Virtual Conferencing: A Participatory Design Study. CyberPsychology <html_ent glyph="@amp;" ascii="&"/> Behavior 4, 287–305 (2001)
[20] Guye-Vuilleme, A., Capin, T.K., Pandzic, S., Thalmann, N.M., Thalmann, D.: Nonverbal communication interface for collaborative virtual environments. Virtual Reality 4, 49–59 (1999)
[21] Bowers, J., Pycock, J., O'Brien, J.: Talk and embodiment in collaborative virtual environments. In: Proceedings of the SIGCHI conference on Human factors in computing systems common ground - CHI 1996, Vancouver, British Columbia, Canada, pp. 58–65 (1996)
[22] Lantz, A.: Meetings in a distributed group of experts: comparing face-to-face, chat and collaborative virtual environments. Behaviour & Information Technology 20, 111–117 (2001)
[23] Allmendinger, K.: Social Presence in Synchronous Virtual Learning Situations: The Role of Nonverbal Signals Displayed by Avatars. Education Pyshcology Review 22, 41–56
[24] Frecon, E., Nou, A.A.: Building distributed virtual environments to support collaborative work. In: Proc. VRST 1998, pp. 105–113 (1998)
[25] Greenhalgh, C., Benford, S.: MASSIVE: a collaborative virtual environment for teleconferencing. ACM Trans. Comput.-Hum. Interact. 2, 239–261 (1995)
[26] Lucia, A.D., Francese, R., Passero, I., Tortora, G.: SLMeeting: supporting collaborative work in Second Life. In: AVI 2008, pp. 301–304 (2008)
[27] Shami, N.S., Cheng, L., Rohall, S., Sempere, A., Patterson, J.: Avatars Meet Meetings: Design Issues in Integrating Avatars in Distributed Corporate Meetings. Group (2010)
[28] Moore, C.: Movement And Making Decisions: The Body-mind Connection. Dance & Movement Press (2005)

Why Can't a Virtual Character Be More Like a Human: A Mixed-Initiative Approach to Believable Agents

Jichen Zhu[1], J. Michael Moshell[1], Santiago Ontañón[2], Elena Erbiceanu[1], and Charles E. Hughes[1]

[1] University of Central Florida, USA
[2] Artificial Intelligence Research Institute (IIIA-CSIC), Spain
jzh@mail.ucf.edu, jm.moshell@cs.ucf.edu, santi.ontanon@gmail.com, eelena@eecs.ucf.edu, ceh@cs.ucf.edu

Abstract. Believable agents have applications in a wide range of human computer interaction-related domains, such as education, training, arts and entertainment. Autonomous characters that behave in a believable manner have the potential to maintain human users' suspense of disbelief and fully engage them in the experience. However, how to construct believable agents, especially in a generalizable and cost effective way, is still an open problem. This paper compares the two common approaches for constructing believable agents — human-driven and artificial intelligence-driven interactive characters — and proposes a mixed-initiative approach in the domain of interactive training systems. Our goal is to provide the user with engaging and effective educational experiences through their interaction with our system.

Keywords: Mixed-initiative system, character believability, interactive storytelling, artificial intelligence, interactive virtual environment.

1 Introduction

Interactive computer characters are becoming widely used in digital systems in a range of domains. Some applications include, for example, virtual patients used in training medical students' communication skills [1], anthropomorphic robotic assistants that can support manual tasks in industrial environments [2], and non-player characters (NPC) in computer games that provide players with dramatic experiences [3]. These virtual or robotic characters are an important part of human centered computing not merely because they provide users with a set of familiar metaphors and conventions to interact with new technology. Like Disney's animated characters [4], they also have the powerful potential to engage users in emotional and personal ways. With a significant improvement of their appearance thanks to recent development of real-time computer graphics technology, there is an increasing demand for these characters to behave in ways that are believable to the users.

However, how to construct believable agents, especially in a generalizable and cost effective way, is still an open problem. This paper compares the two common approaches

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 289–296, 2011.

for constructing believable agents — human-driven and artificial intelligence-driven — and proposes a mixed-initiative approach in the domain of interactive training systems. Our goal is to provide the users with engaging and effective educational experiences through their interaction with our system.

This paper is organized as follows. Section 2 provides a background of two primary approaches for creating believable interactive characters – the human-driven and computer-driven characters. Next, Section 3 discusses the need of combining the strength of the previous approaches and proposes a mixed-initiative system in the domain of a specific interactive education system.

2 Background

This section describes two major existing approaches for creating believable, human-like virtual characters. We mainly draw from the puppetry tradition in the performing arts and the autonomous agent approach from the artificial intelligence (AI) community.

2.1 Human-Driven Puppetry

Performing arts and the entertainment industry have a long-standing tradition of creating believable and engaging characters. In theater, for instance, human actors have developed various strategies to enact characters, either with their own body or through inanimate props (e.g. puppets). Grown out of this tradition are digital puppetry systems, which allow human actors to sit behind the curtain and "Wizard of Oz" virtual characters.

Often used in theme parks and live theatre, Digital puppetry relies mainly on human operators' skills to control a substantial part of virtual characters' movement and/or speech in order to create an illusion of life. For instance, Disney's tradition of emotional involvement with the audience using their classic animated characters continues with its modern "Turtle Talk with Crush" show, where performers puppeteer the virtual turtle character in a 15-minute improvisational interaction and conversation with a large audience. More recently, this approached is further developed in the emerging machinima community, where player-controlled game characters enact cinematographic scenes in real-time game engines.

2.2 Autonomous Believable Agents

A second broad approach for constructing human-like virtual characters is developed by the Artificial Intelligence (AI) community. AI researchers have spent decades constructing autonomous agents capable of accomplishing a variety of tasks at human-level intelligence. Among them, Joe Bates and his students in the OZ project at CMU [5] coined the term "believable agents" to describe what many consider to be the "holy grail" of computer-driven characters in the context of interactive drama. These computer agents utilize various AI techniques to create the illusion of personality, emotion, and lifelikeness. Since then, the notion of believable agents has been explored in various domains such as interactive narrative, story generation systems, software agents, and robotics. A salient example is the believable agents (i.e. Trip and Grace) in Mateas and Stern's acclaimed interactive drama Façade [6].

In comparison, human-driven (digital) believable agents are flexible and more robust in difficult scenarios such as free conversation in public spaces. However, their level of believability is strongly tied to the expertise of the human operator, and this approach can be very costly. Its AI-driven counterpart, on the other hand, has lower run-time cost and can, in theory, operate continuously. But this second approach requires substantial authorial burden in advance and, more importantly, many of the needed technological components (e.g., perception, common-sense reasoning, and natural language processing) are still far from perfect. As a result, most autonomous believable agents only function properly in very constrained environments.

3 Building Believable Agents in Educational Systems

In order to combine to strength of both approaches described above, this section proposes a mixed-initiative framework for creating believable characters. Below, we will briefly describe our application domain, the TeachME™ system, and propose a mixed-initiative system to improve the synergistic collaboration between the human inter-actors and the system.

3.1 A Mixed-Initiative Approach

Mixed-initiative systems have received increasing attention in domains such as robotics [7], but are relatively under-explored in virtual characters [8, 9]. Our goal is to provide users with engaging and effective educational experience by creating believable agents that exhibit comparable level of believability as human or human-controlled characters with relatively low requirements/costs. For example, we would like to provide a relatively large number of believable agents per moderately-experienced human operators.

From the human computer interaction (HCI) point of view, mixed-initiative interfaces are situated between the two major interface paradigms — direct manipulation (e.g. a word processor) and intelligent agent (e.g. automated web bots) [10]. Typically applied to systems with a mixture of human agents and computer agents, the term "mixed-initiative" refers to a flexible control strategy where each agent can contribute to the task that it does best by temporarily taking control of the task (i.e. having initiative) [11]. Seen as an "elegant coupling" of automated services with a user's ability to directly manipulate interfaces to access information [12], mixed-initiative systems are gaining popularity in both software and robotic systems. A growing belief is that the collaborative control leads to a more balanced and symbiotic relation between human and computer/robotic agents based on their respective capabilities. By coordinating both parties to request and receive help on actions that they could not have performed alone, mixed-initiative systems can decrease the need for continuous human monitoring and increase overall performance [7, 13].

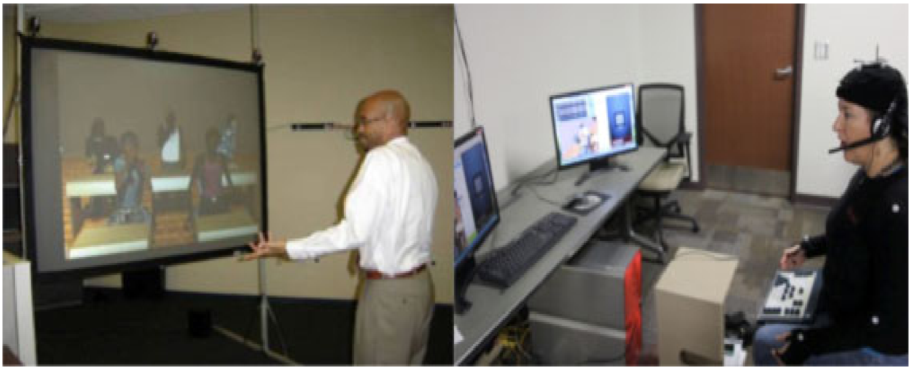


Fig. 1. The TeachME™ System (Left: User's View; Right: Inter-actor's View)

3.2 The TeachME™ Project

Over the past four years, the lab directed by Hughes has developed several systems in which human inter-actors "puppeteer" virtual characters in interactive systems designed for training and educational purposes. This section gives a brief account of the TeachME™ project, which will be used as the specific domain for us to test our research on mixed-initiative believable characters.

Orlando with its numerous theme parks and entertainment venues is home to a large inter-actor community. In 2007, under the direction of Hughes, a software team led by Dan Mapes of the Institute for Simulation and Training (IST) collaborated with a group of inter-actors led by Jeff Wirth and a team of educators led by Lisa Dieker and Mike Hynes from the College of Education at the University of Central Florida (UCF). They developed a teacher training system called TeachME™ (Teaching in a Mixed Reality Environment). The system provides pre-service and in-service teachers the opportunity to learn teaching skills and to craft their practice without placing "real" students at risk during the learning process. The original application was classroom management [14]. Since then, the system has been used in mathematics instruction, writing classes and counselor training. Starting in summer 2009, Utah State University adopted the system for working with its teachers of children with special needs. Three new partnerships formed in fall 2010 with Old Dominion University, West Virginia University and University Center of Greenville South Carolina; Florida State University, University of Wisconsin Milwaukee and Western Michigan University joined in spring 2011; and there are now emerging partnerships with at least five other teacher preparation programs.

In TeachME™, the user (i.e. teacher trainee) can interact with 3D animated virtual students in a middle-school classroom setting, which is displayed at life size on a rear projection screen (Figure 1, left frame). Through natural language conversation, gesture and body position, the user can interact with the virtual characters. For instance, she can walk toward one of the five virtual students (to a marked spot on the floor) and ask him a question. Using simple two-camera tracking, the system automatically shifts the virtual camera to zoom in on this simulated student.

A group of five virtual students are controlled by an inter-actor behind the scenes (Figure 1, right view). We use infrared-marker motion capture of the inter-actor's head and upper body (the virtual students are seated at desks). We transmit raw position/orientation data over the Internet, along with audio/video data over Skype. Each student's behavior is based on one of the personality types identified by Long [15] and Dreikurs [16]. The team of inter-actors developed detailed back-stories and unique personalities for each student.

Unlike many AI-driven agents that rely on a small pool of pre-recorded voice responses, the characters in TeachME™ behave and converse in a much more believable and human way; after all, a live inter-actor is controlling any virtual student to whom the user directly communicates. Even as the user focuses on one student, the other students carry on their standard pre-programmed behaviors. If the teacher ignores the virtual students as individuals or shows little understanding of their personal motivations, they will get more rowdy or more withdrawn (depending on the character's attributes). Such increases in the level of chaos can be controlled individually or across the entire class by the inter-actor (primarily) or triggered by an on-site trainer through a simple key-based system. This project gives life to the virtual classroom experience, causing many trainees to feel that they spent a half hour in instruction, when the exercises typically last only five minutes.

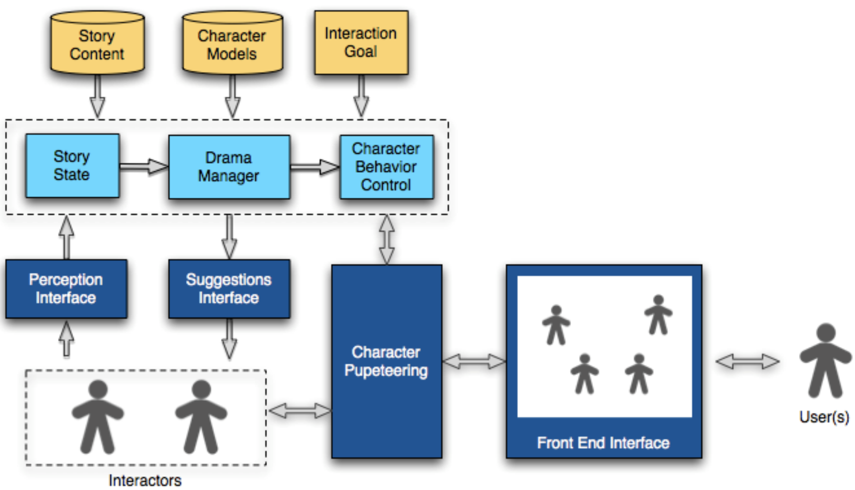


Fig. 2. Proposed Mixed-Initiative System Architecture

3.3 Proposed System

One key element to achieve our goal of improving the efficiency of creating believable characters is to alleviate the cognitive load of the human operators during the performance and distribute the tasks between the humans and the computer in ways which enhance both parties' strengths. Based on our preliminary study of the human operators in the TeachME™ project, we classified the tasks involved in constructing believable characters into human-friendly and computer-friendly ones. For instance, human are particularly effective in handling freestyle natural language conversations with the user and observing the user's physical and psychological reactions, all of which are widely known as difficult tasks for AI. The computer, on the other hand, is more efficient in routinizable tasks, such as idle animations of the characters. It is also good at overseeing the over-arching experience and adapting the structure of the story based on user interaction.

Different from both of our existing digital puppetry systems and Niehaus and Weyhrauch's proposed mixed-initiative system [17], this proposed system will have an explicit representation of the story. It will allow the overall system (including both the human inter-actors and the AI component) to handle much more complicated narrative structures where the storyline can change algorithmically based on user interaction and allow more user agency in the learning experience.

We loosely divide the system's operation into two main tasks: enacting the believable characters and maintaining the storyline at the plot level. The first task includes 1a) providing characters' direct interactions with the user (e.g. gesture and dialogue), 1b) maintaining the characters' higher-level consistency (e.g. personalities, emotions, patterns of behavior, background story, etc.). The second main task includes 2a) instantiating the pre-defined plot structure based on user interaction, and 2b) restructuring the plot if the initial story fails to work; e.g., if the user continually fails a specific learning task, the story should be reorganized to emphasize similar tasks. Both tasks require 3) the perception of user interaction and 4) the coordination between inter-actor and the AI component to resolve any major potential conflict.

Based on our initial observation described above, a human inter-actor is particularly capable of task 1a), for one character at any point of time, and 3). Both tasks are notoriously difficult for computer systems to accomplish. On-going research in interactive narrative/computer games [6, 18] and story generation [19, 20] has shown that the state of the art artificial intelligence techniques are capable of handling tasks 2a), partially 2b). In this iteration of our system design, the inter-actor and the computer will work together on tasks 1b), 2b) and 4).

Figure 2 above shows our design for the proposed system. The AI-based Experience Management System (EMS) consists of the states of the characters and the story, a drama management module, and a character behavior control module. The drama management module is responsible for executing and reconstructing the events on the fly so that they will best satisfy the overarching goal. Generally speaking, the character behavior control module can control the characters that are not currently controlled by the inter-actor. EMS also contains computational models of pre-authored characters, story content, and the overall goal of the whole interactive experience.

The main human inter-actor provides one character's direct interaction with the user and can partially control the actions of other characters that she is not directly enacting. The character that she enacts can shift in time. A second inter-actor is primarily responsible for handling perception for EMS.

In our system architecture, EMS and the inter-actors communicate mainly in two ways. EMS provides high-level information and suggestions to the main inter-actor. For instance, it will display the emotional state for each character, and provide possible conversation topics and background stories to the main inter-actor. The second inter-actor will observe the actions and reactions of the user and input such information to the computer in real time. As computer vision techniques become more sophisticated, the perception task may be partially handled by the computer in the future.

4 Conclusion

This paper proposes a mixed-initiative system to create believable agents in interactive educational systems. The key challenges for our approach are: recognition of different capacities of the human and the computer; design of the mixed-initiative control scheme that is intuitive to human operators; construction of believable characters with relatively low cost; and achievement of the overarching educational goal. Some of these challenges are of particular relevance to HCI because they call for new interaction models which facilitate the shift of shared control between the human and the computer.

References

1. Stevens, A., Hernandez, J., Johnsen, K., Dickerson, R., Raijb, A., Harrison, C., DiPietro, M., Allen, B., Ferdig, R., Foti, S., Jackson, J., Shin, M., Cendan, J., Watson, R., Duerson, M., Lok, B., Cohen, M., Wagner, P., Lind, D.S.: The use of virtual patients to teach medical students history taking and communication skills. The American Journal of Surgery 191, 806–811 (2006)
2. Helms, E., Schraft, R.D., Hagele, M.: rob@work: Robot assistant in industrial environments. In: Proceedings of the 11th IEEE International Workshop on Robot and Human Interactive Communication, pp. 399–404 (2002)
3. Mateas, M.: Interactive Drama, Art, and Artificial Intelligence Ph.D. Dissertation, Computer Science, Carnegie Mellon University, Pittsburgh PA (2002)
4. Thomas, F., Johnston, O.: The Illusion of Life: Disney Animation. Walt Disney Productions (1981)
5. Bates, J.: The role of emotion in believable agents. Communications of the ACM 37, 122–125 (1994)
6. Mateas, M., Stern, A.: A Behavior Language for Story-Based Believable Agents. IEEE Intelligent Systems 17, 39–47 (2002)
7. Fong, T., Thorpe, C., Baur, C.: Robot as Partner: Vehicle Teleoperation with Collaborative Control. In: Proceedings from the 2002 NRL Workshop on Multi-Robot Systems (2002)
8. Goodrich, M.A., Olsen Jr., D.R., Crandall, J., Palmer, T.J.: Experiments in Adjustable Autonomy. In: Proceedings of IJCAI Workshop on AUtonomy, Delegation and Control: Interacting with Intelligent Agents, pp. 1624–1629 (2001)
9. Harrell, D.F., Zhu, J.: Agency Play: Dimensions of Agency for Interactive Narrative Design. In: Proceedings of AAAI 2009 Spring Symposium on Intelligent Narrative Technologies II, Stanford, CA, pp. 44–52 (2009)
10. Shneiderman, B., Maes, P.: Direct Manipulation vs. Interface Agents: A Debate. Interactions IV, 42–61 (1997)
11. Hearst, M.A.: Mixed-initiative Interaction. IEEE Intelligent systems 14, 14–24 (1999)

12. Horvitz, E.: Principles of Mixed-Initiative User Interfaces. In: Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 1999), Pittsburgh, pp. 159–166 (1999)
13. Rosenthal, S., Veloso, M.: Using Symbiotic Relationships with Humans to Help Robots Overcome Limitations. In: Workshop for Collaborative Human/AI Control for Interactive Experiences, Autonomous Agents and MultiAgent Systems, AAMAS (2010)
14. Dieker, L., Hynes, M., Hughes, C.E., Smith, E.: Implications of Mixed Reality and Simulation Technologies on Special Education and Teacher Preparation. Focus on Exceptional Children 40, 1–20 (2008)
15. Long, S.: Malevolent Estrangement. Youth & Society 7, 99–129 (1975)
16. Dreikurs, R., Grunwald, B.B., Pepper, F.C.: Maintaining sanity in the classroom. Harper and Row, New York (1968)
17. Niehaus, J., Weyhrauch, P.: An Architecture for Collaborative Human/AI Control of Interactive Characters. In: Workshop for Collaborative Human/AI Control for Interactive Experiences, Autonomous Agents and MultiAgent Systems, AAMAS (2010)
18. Harrell, D.F.: GRIOT's Tales of Haints and Seraphs: A Computational Narrative Generation System. In: Wardrip-Fruin, N., Harrigan, P. (eds.) Second Person: Role-Playing and Story in Games and Playable Media. MIT Press, Cambridge (2007)
19. Meehan, J.: The Metanovel: Writing Stories by Computer, Ph.D. Dissertation, Computer Science Department, Yale University, New Haven (1976)
20. Riedl, M.O., León, C.: Toward Vignette-Based Story Generation for Drama Management Systems. In: Proceedings of the 2nd International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN), Workshop on Integrating Technologies for Interactive Story, Cancun Mexico (2008)

Part IV
Developing Virtual and Mixed Environments

Collaborative Mixed-Reality Platform for the Design Assessment of Cars Interior

Giandomenico Caruso, Samuele Polistina, Monica Bordegoni, and Marcello Aliverti

Dipartimento di Meccanica, Politecnico di Milano, Via La Masa, 1
20156 Milano, Italy
{giandomenico.caruso,monica.bordegoni}@polimi.it,
samuele.polistina@mecc.polimi.it, marcello.aliverti@kaemart.it

Abstract. The paper describes a collaborative platform to support the development and the evaluation of cars interior by using a Mixed Prototyping (MP) approach. The platform consists of two different systems: the 3D Haptic Modeler (3DHM) and the Mixed Reality Seating Buck (MRSB). The 3DHM is a workbench that allows us to modify the 3D model of a car dashboard by using a haptic device, while the MRSB is a configurable structure that enables us to simulate different driving seats. The two systems allow the collaboration among designers, engineers and end users in order to get, as final result, a concept design of the product that satisfies both design constraints and final users' preferences. The platform has been evaluated by means of several testing sessions, based on two different scenarios, so as to demonstrate the benefits and the potentials of our approach.

Keywords: Collaborative design, Mixed Reality, Virtual Prototype, Haptic modeling, Ergonomic assessment.

1 Introduction

The development process of cars interior includes the execution of several evaluation tests necessary for the optimization of the final product. Specifically, the feedback acquired during the evaluation tests performed with end users will determine the commercial success of the product. The data concerning the evaluation with end users are generally acquired at the final phases of product development, since a physical prototype of the product is necessary for the execution of such kind of tests. Consequently, these data cannot be used for deeply modifying the shape of the car interior and, they will be used only for design restyling.

Automotive industries are strongly interested in performing evaluation tests early in the design phase. These tests involve different experts, who have to collaborate for defining a solution that simultaneously satisfies design constraints and end users' preferences. Interior car design, in fact, is not focused only on aesthetic values and safety requirements, but it is also strongly related to ergonomic aspects, which obviously require the study of interaction between the user and the dashboard components.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 299–308, 2011.

Our research has aimed at developing a collaborative platform based on Mixed Reality (MR) technologies that enables designers to assess and refine the shape of a car interior in a more natural and interactive way while enabling other experts (engineers and ergonomists) to perform some evaluation tests with end users with the objective of getting users' acceptance without the need to build several physical prototypes, and thus reducing the product development time.

This paper is structured as follows: section 2 presents an overview of related works, in section 3 we describe the collaborative platform architecture and the hardware and software implementation. Section 4 reports about the usability of the platform assessed by performing different tests related to two different scenarios deriving from typical ergonomic issues of car interiors. Finally, in section 5 and 6 we discuss some considerations and present our conclusions.

2 Related Works

It has become accepted practice the use of the term "collaborative systems" for describing the computer systems that support distal communication between designers [1]. The design activities, in fact, involve several professional figures, who have to collaborate for defining the final product. These activities range from modeling to numerical simulations, evaluations with end users and so on. Therefore, the issues related to the collaborative environment are very complex and several authors propose different approaches to overcome them. Madsen [2] identifies some of the common collaborative barriers such as language differences, time zones, miscommunication, ambiguity in requirements, misunderstandings of design intent and proposes a collaborative strategy to overcome these barriers in particular for supporting distributed teams in CAD systems integration. Li et al. [3], highlight that, due to the complexity of collaborative design activities, a collaborative system cannot be a simply set-up obtained through equipping a stand-alone CAD system with IT and communication facilities, but it needs several innovations or even fundamental changes such as infrastructure design, communication algorithms, geometric computing algorithms, etc. For this reason, the role of VR technologies [4] and in particular of Virtual Prototypes (VPs) [5] is fundamental in this context and several examples demonstrate their effectiveness during the product design process [6-12]. However, the use of VPs arise many issues related to the interaction, in particular during the modification or the evaluation phases of new products. MR technologies can enhance the interaction with the VP by providing a more realistic visualization and by adding the haptic feedback. For instance, Mixed Prototyping (MP) approach, which consists in creating a prototype partially real and partially virtual can be effectively used for the rapid design assessment of new products, as described in [13].

The evaluation of car driving seats with end users, instead, needs systems, named seating bucks, which simulate the car interiors. Many research groups and industrial research centers use this kind of systems. The ELASIS research group [14], for example, has developed a system based on a parametric driver's seat simulator, which supports automatic configuration of steering wheel, driver seat and pedals, coupled with an immersive VR environment [15]. H. Salzmann et al. developed a two-users virtual seating buck system [16], which enables two users to take the function of the driver and co-driver respectively. However, these seating bucks are developed by using VR immersive environments that reduce the perceived realism of the scene: in fact, several users, especially when they have to wear a Head Mounted Display, complain about an unnatural perception of space. Also in this activity, MR technologies can improve the users' interaction with these systems. Ohshima et al. [17], for instance, describe how a MR system can improve the perception of the distances than when the only the visual sense is available, as happens in the VR environments. We have also proposed, in a recent work [18], an innovative MR seating buck system, which enables us to easily compare, with end users, different car interiors.

The above-presented overview has shown several systems for supporting collaborative design and the evaluation of cars interior; however, none of these systems enables performing these two different activities collaboratively. Our work intends to demonstrate how MR technologies can support these activities by creating a collaborative platform based on Mixed Prototyping approach.

3 The Collaborative Platform

The main purpose of our work is to provide an effective system for easily and interactively modifying CAD models of cars interior in real scale and for verifying in real time the modifications with the end users. Usually, these two activities are carried out in two different phases of the car interior development and are often performed by two teams working in two different locations. The Collaborative Platform aims at fostering the concurrent performance of these two activities in this particular situation. For this reason, we set the two systems in two different locations during the testing sessions and the participants were able to communicate by using commercial video-conference software. The designer can modify the 3D model through the 3DHM system while other experts assess the goodness of the modifications by means of the MRSB with an end user. If the user is not satisfied with the results of the modification, the designer is able to further modify the model, according to the comments of the user and of the other experts. Subsequently, the modified model is proposed again to the user, who can express a new judgment. The evaluation test is considered finished when all the participants judge positively the final model.

The Collaborative Platform mainly consists of two different systems: the 3D Haptic Modeler and the Mixed Reality Seating Buck.

(a) (b)

Fig. 1. The Mixed Reality Collaborative Platform: a) 3D Haptic Modeler, b) Mixed Reality Seating Buck

The *3D Haptic Modeler* (3DHM) is a workbench integrating stereoscopic visualization and haptic interaction (Fig.1a). The display system consists in an original solution that we have named Direct Visuo-Haptic Display System (DVHDS), based on mirrors and screens for projecting a 3D image over the haptic workspace [19]. Such solution allows us to visualize in real scale the area of a car interior, which has to be modified. For the haptic interaction, we have used the MOOG-HapticMaster (HM) haptic device [20]. This device is able to render very high rigidity with low friction and, above all, its workspace is large enough to guarantee an appropriate working space. Using a haptic device with a small workspace, as for example the Sensable Phantom device [21], would not allow covering the whole area of a car interior visualized in real scale. Unfortunately, the HM device does not allow users to haptically render any kind of surface. The HM, in fact, can only render haptic effects, such as dampers and springs, and spatial geometrical primitives can be defined, such as spheres, cones and cubes [22]. In order to solve this problem, we have implemented an original algorithm.

The *Mixed Reality Seating Buck* (MRSB) is a system that allows us to simulate different driving seat set-ups for performing different tests enabling the users to see and interact in a natural way with the Mixed Prototype of a vehicle (Fig 1b). The MRSB mainly consists of a configurable hardware structure, which simulates the driving seat, an Optical See-Through Head-Mounted Display (OST-HMD) that enables the user to visualize the virtual representation of the car interior and the real environment at the same time, and a robotic arm, which allows us to interactively change the position of some physical components of the dashboard (for example, buttons, sliders, and knobs). Finally, an optical tracking system equipped with 6 cameras tracks some elements of the structure, the user's point of view and the user's hands for aligning virtual components to the real ones.

The software application, developed for the collaborative platform, consists of five modules: the MainLoop module, the OpenSG module, the ThinkCore module, the HapticMaster module, and the Kuka module, as shown in Fig. 2.

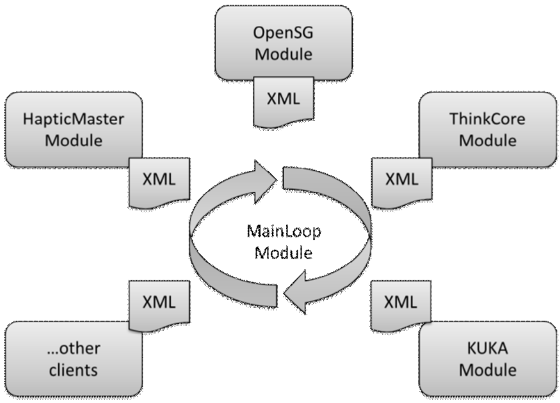


Fig. 2. Software architecture

The core of the system is the MainLoop module. It is based on a publish/subscribe paradigm for interprocess communication based on XML messages sent over a TCP/IP connection [23]. Thanks to that, the 3DHM and the MRSB systems can be placed in different locations and easily connected by Ethernet communication. The MainLoop manages the exchange of data within the other modules of the platform and permits the definition of a specific behavior for each module using a finite-state machine approach. The MainLoop module is also able to connect other clients, which extend the platform in order to provide other workstations in different locations.

The OpenSG module manages the visualization both of the 3DHM and of the MRSB through two stereo viewports that support the active stereo and the passive stereo modalities. OpenSG is an open source portable scene-graph system able to create real-time graphics programs by supporting all the modern computer graphics features.

The ThinkCore module, instead, is based on a subset of the think3 COM API [24]. This module allows us to use the functionalities of ThinkDesign CAD system with the aim of developing applications working with CAD models. The ThinkCore module manages the CAD model by generating a tessellated representation of the model and by providing the haptic objects stored in the model. The tessellated representation is sent, through the MainLoop module, to the OpenSG module for the stereoscopic visualization while the haptic objects are sent to the HapticMaster module for the haptic interaction. The haptic modifications, in fact, are made directly on the CAD model while the ThinkCore module deals with the updating of the visualization. In addition, this module provides us with two powerful CAD modification modalities of ThinkDesign: the Interactive Solid Modeling (ISM) and the Global Shape Modeling (GSM). Such modification modalities enable us to directly modify the CAD model by using the haptic interaction modalities, which are simple and intuitive.

The HapticMaster module manages the haptic behavior of the HM devices and enables us to acquire data from the end-effector position and to generate the objects used for the haptic integration. Since the HM is able to interact only with simple primitives, we have implemented a procedure that enables the user to touch the general surfaces of the CAD model. The procedure creates a Haptic Virtual Plane according to the end-effector position with the same normal of the selected surface. The virtual plane moves according to the end-effector position, while the ThinkCore module elaborates the new normal in respect to the surface in real time, and gives the user the feel to touch the surface. This tangent plane is determined by a normally oriented spring that is calculated in the position of the HM end effector. Finally, through the HapticMaster module, we have created a sort of magnetic point, named Snap Point (SP), which helps the user to haptically select a specific feature of the CAD model. We had to implement such haptic object because similar objects do not exist in the HM API. The SP is a haptic object made up of three haptic springs with the same application point but different stiffness and deadband.

The Kuka module manages the communication with the robotic arm. In particular, the module sends to the robot the position that has to be reached by its end effector. These positions are obtained directly from the modification made on the CAD model.

4 Evaluation Tests

The case studies, which we have chosen for the evaluation test, have been carried out by using two typical ergonomic issues of car interiors. The first one consists in the evaluation of the reachability relating to the position of the climate control system knob on the car dashboard while the latter one is based on the visibility assessment of the left A-pillar. In the automotive field, such issues are considered very useful to improve the quality of the final product and the level of satisfaction for costumers.

The repositioning of the knob has been made by using the ISM modification modality. The expert user moves one of the three knobs to a new position according to his preferences by using the 3DHM. During the modification, the haptic feedback link the knob to the surface of the dashboard and a “ghost” representation of the knob help the user to define the final result of the modification. Simultaneously, the new geometry is sent to the MRSB and the physical prototype of the knob is moved by the robot in the new position. Therefore, the user seated on the MRSB can evaluate the new position of the knob both visually, by means of OST-HMD, and haptically, by touching the real knob. Fig. 3 shows the modification of the model during the knob repositioning.

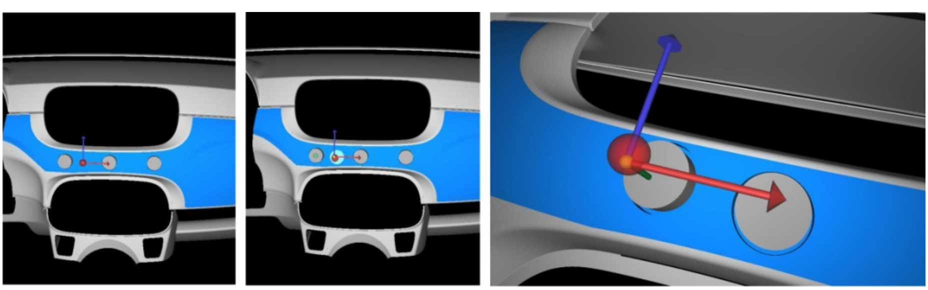


Fig. 3. ISM modification applied during the knobs repositioning

The second case study, instead, concerns the A-pillar restyling and the modification is done by using the GSM modification modality. During the modification, the haptic feedback simulates the surface deformation and a graphical representation of the result is updated in real time. These feedbacks help the user to define the final result of the modification. At the end of the modification phase the user seated on the MRSB can evaluate visually the goodness of the modification. The test ends when the user is satisfied with the achieved A-pillar modifications. Fig. 4 shows the main steps of the dashboard restyling.

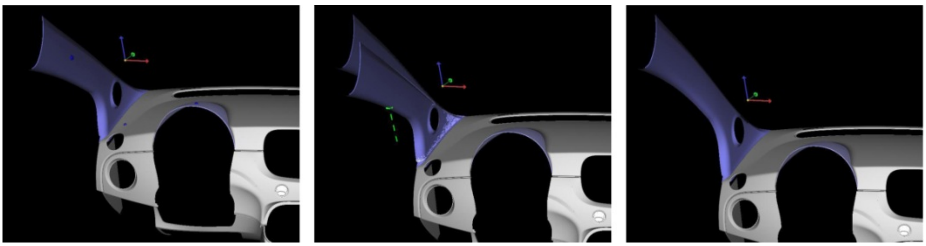


Fig. 4. GSM modification applied during the visibility test of the A-pillar

5 Discussion

In this section we discuss the data collected during the testing sessions in reference to the usability aspects of the collaborative platform. The definition of a protocol for evaluating the effectiveness of our platform is necessary and we have investigated several approaches in order to define the most suitable one. We have decided to adopt the heuristic method proposed by Nielsen et al. [25] that enables us to address the usability issues of the collaborative platform. In particular, during the test we have involved an expert user for assessing the usability of the 3DHM and other 10 users, who are skilled in VR technologies, for the evaluation of the MRSB functionalities. The expert user has a background on heuristic methodology, with 3-years experience in HCI research, so we can classify him as an HCI expert user. The expert user and the other users have been invited to carry out some specific tasks according to the two case studies, which will be described in the following. At the end of the testing session, all the participants have filled a questionnaire that has been elaborated according to the Nielsen's heuristics. Table 1 presents the correlation between the data collected during the testing sessions and the qualitative usability aspects investigated. Some of such data have been collected during the execution of the testing sessions while the remaining ones are the users' judgments, related to some aspects of our collaborative platform, expressed on a scale from 1 (bad) to 10 (good). In this table, we correlated quantitative data collected and qualitative aspects for assessing the usability of our system.

Table 1. Usability Considerations

Measurements	Value	Usability aspects
Misunderstanding	3%	Learnability
System errors	2%	Margin of error
User's errors	10%	
Influence of surrounding environment	8	Efficiency
Influence of the structure	7	
Field of view limitation	4	
Time for tasks execution	9	
Perceived Comfort	6	Satisfaction
Perceived Realism	4	
Global evaluation	6	

At first, we correlated the system learnability of the platform to the times that the user misunderstood some task during the testing session and asked to explain it. This datum is presented as the rate between the number of misunderstandings and the total number of tasks (30 for each testing session) and highlights a good affordance of the platform. Obviously, this datum will be more significant when the sample of users will be wider and more heterogeneous since, in this testing session, the involved users are skilled on VR technologies.

The margin of error is an aspect that we correlated with the errors occurred during the execution of the testing session. We divided such data into two categories: system and user's errors. The first one represents the times that one of the components constituting the system (robotic arm, tracking system, haptic device, etc.) goes in failure mode during the test. In the second category similar kind of errors are collected but only if they occurred when the user interacts with the system. These values are expressed as the rate between the number of errors and the number of tasks in which we subdivided the testing session. We considered the margin of error one of the aspects that can represent the reliability of our system.

The efficiency of the system, instead, has been correlated to users' judgments that relate to some issues, which can limit the interaction and consequently invalidate the assessment. The surrounding environment and the presence of the structure, for instance, can influence negatively the user during the testing session since they are visual noises. The time for task execution is another issue investigated for assessing the efficiency and the effectiveness of our approach. Thanks to the robotic arm, the time needed for changing the layout is very short (few seconds) and thereby the users are able to correctly compare different proposed solutions and the time need to complete all the testing session (about 20 minutes) did not wearied any user.

Globally, the users' judgments have been positive also in relation to the aspects related to the satisfaction in using the system. These results are encouraging and show the effectiveness of our platform that certainly improves the normal activities carried out for evaluating car interiors. Unfortunately, we cannot do a comparative assessment since now there are not similar collaborative platform to compare. However, the traditional procedure, which the automotive industries follow to validate a car interior, is complex and implies a lot of downtime.

6 Conclusion

This paper has presented our MR collaborative platform for the design assessment of the car interiors. The aim of this platform is to improve the decision-make process during the project development by providing the possibility of verifying the design modifications in real time with the end users. The conducted testing sessions confirm the good usability of the platform. However, some technological issues have to be solved for improving the effectiveness of our collaborative platform. In the next future, we aim at improving the system as regards the interface for the user's interaction, and at solving the technological issues arose during the testing sessions.

Acknowledgments. The research work presented in this paper has been supported by the Italian Ministry of University and Research in the frame of the FIRB-PROGIMM project (http://www.kaemart.it/progimm). The authors would like to thank Sabrina Magni for her precious contribution.

References

1. Kvan, T.: Collaborative design: what is it? Automation in Construction 9(4), 409–415 (2000)
2. Madsen, K.: Collaboration strategies for distributed teams: A case study of cad systems integration. In: Fourth International Conference on Systems, pp. 222–227 (2009)
3. Li, W.D., Lu, W.F., Fuh, J.Y.H., Wong, Y.S.: Collaborative computer-aided design-research and development status. Computer-Aided Design 37, 931–940 (2005)
4. Jimeno, A., Puerta, A.: State of the art of the virtual reality applied to design and manufacturing processes. The International Journal of Advanced Manufacturing Technology 33(9), 866–874 (2007)
5. Yuan, Q.K., Guo, C.R., Li, S.J., Liu, D.H.: A products integrated & collaborative design methodology based on virtual product prototypes. In: Proc. of Information and Automation, pp. 1023–1027 (2009)
6. Li, S.K., Guo, Y., Yang, Q., Zhao, W.H.: Collaborative virtual prototyping and collaborative design methodology. Journal of System Simulation 13(1), 124–127 (2001)
7. Song, X., Li, B.H., Chai, X.D.: Research of complex product virtual prototype lifecycle management system. Journal of System Simulation 19(11), 2484–2487 (2007)
8. Wang, P., Li, B.H., Chai, X.D., Yan, X.F.: Collaborative modeling technology of virtual prototype for complex product. Journal of System Simulation 16(2), 274–281 (2004)
9. Li, J., Zhi, G.W., Oraifige, I.A.: Distributed VR for collaborative design and manufacturing. In: 11th International Conference on Information Visualization, pp. 744–753 (2007)
10. Shen, W.M., Barthes, J.P.A.: Special Issue on collaborative design and manufacturing. Advanced Engineering Informatics 22(3), 281–286 (2008)
11. Zeiler, W., Quanjel, E.: Integral design methodology within industrial collaboration. In: Proc. of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 633–641 (2008)
12. Chen, Z.Q., Siddique, Z.: A distributed environment for collaborative design. In: Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 759–768 (2008)

13. Bordegoni, M., Cugini, U., Caruso, G., Polistina, S.: Mixed prototyping for product assessment: a reference framework. International Journal on Interactive Design and Manufacturing 3(3), 177–187 (2009)
14. ELASIS, http://www.elasis.it
15. Caputo, F., Gironimo, G.D., Monacelli, G., Sessa, F.: The design of a virtual environment for ergonomic studies. In: Proc. of ADM International Conference, Rimini (2001)
16. Salzmann, H., Froehlich, B.: The Two-User Seating Buck: Enabling Face-to-Face Discussions of Novel Car Interface Concepts. In: Proc. of IEEE Virtual Reality Conference, pp. 75–82 (2008)
17. Ohshima, T., Kuroki, Y.T., Hiroyuki, T.H.: A Mixed Reality System with Visual and Tangible Interaction Capability - Application to Evaluating Automobile Interior Design. In: Proc. of 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality (2003)
18. Caruso, G., Tedioli, L.: Mixed reality seating buck system for ergonomic evaluation. In: Proc. of TMCE 2010 Symposium, Ancona, Italy (2010)
19. Bordegoni, M., Covarrubias, M.: Direct Visuo-Haptic Display System Using a Novel Concept. In: 3th Eurographics Symposium on Virtual Environments (2007)
20. HapicMaster, http://www.moog.com/products/haptics-robotics
21. Phantom, http://www.sensable.com
22. Lammertse, P., Frederiksen, E., Ruiter, B.: The HapticMaster, a new high-performance haptic interface. In: Proc. of Eurohaptics, Edinburgh, UK (2002)
23. Antolini, M., Covarrubias, M., Bordegoni, M., Cugini, U.: A framework for managing multiprocess applications based on distributed finite-state machine approach. In: Proc. of IEEE Ro-MAN (2010)
24. think3, http://www.think3.com
25. Nielsen, J.: Usability inspection methods. In: Proc. of Conference Companion on Human Factors in Computing Systems, pp. 413–414. ACM, New York (1994)

Active Location Tracking for Projected Reality Using Wiimotes

Siam Charoenseang and Nemin Suksen

Institute of Field Robotics, King Mongkut's University of Technology Thonburi,
126 Pracha-u-thit Bangmod Tungkru 10140 Bangkok, Thailand
siam@fibo.kmutt.ac.th, naymin_suksen@msn.com

Abstract. Some addressed issues in projected reality are location acquisition, limited work space, and geometric distortion. This paper proposes a low-cost, robust, fast, and simple method for handling addressed problems using infrared camera in Nintendo's Wiimotes and a pan-tilt camera head. Two Wiimotes are attached on both horizontal and vertical axes of a portable projector mounted on a pan-tilt camera head. Hence, it can detect 4 infrared LEDs on the corners of a display surface in perspective projection volume. The augmented images are wrapped to fit the display area. To increase the system workspace, a pan-tilt camera head is used to track the display surface. While the display surface or the projector moves, a proposed fast location tracking algorithm between two Wiimotes is implemented. Experimental results demonstrate the ability of real time location tracking at 97 fps that is more than the refresh rate of typical projector. Finally, the active location tracking using the pan-tilt camera head can give workspace more than 36 times of the normal perspective projection workspace.

Keywords: Perspective Location Tracking, Projected Reality, Augmented Reality.

1 Introduction

In modern augmented reality application, projected reality approaches are not limit to represent augmentations information. It can enhance human vision sense by changing the texture on surface such as projection surfs on the fall, projection of moving wheel on a static plate. Existing research works in the field of computer vision, human computer interface, augmented reality communities introduce tracking methods for finding projected location such as rotating mirror with video camera [1], attach sensor with magnetic or optical tracker [2], infrared-reflector markers with camera [3], passive marker with rotatable camera [4], 3D surface modeling with a ray/triangle intersection algorithm [5], and structure light patterns with optical fibers [6, 7]. While far from a complete list, those methods manifest the value of flexible image projection to any surface rather than traditional use of projector on passive flat screen.

In this paper, the potential of solution of using the perspective projection of input/output geometry relationship between camera and projector is explored as shown in Figure 1. In an ideal case, camera and projector have the same horizontal

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 309–317, 2011.

and vertical field-of-view (FOV), and alignment. This yields equality between the location of targeted object on camera's plane and projection's plane. However, it is impossible to place camera and projector at the same pose. Hence, this paper proposes a method to make virtual camera in ideal case by using two Nintendo's Wiimotes placing closely to the projector on horizontal and vertical axes.

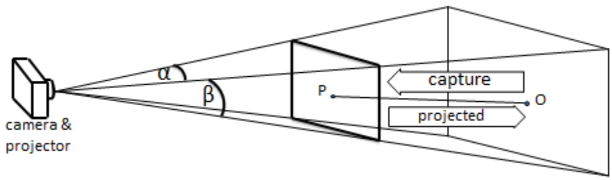


Fig. 1. Ideal case, camera and projector are placed at the same pose. α and β are field of view horizontal and vertical angles, respectively. P is a coordinate of (x,y) on camera and projector planes. O is a coordinate of (x,y,z) on the world space.

2 System Overview

The purposed location tracking system includes a portable projector, four diffusion infrared LEDs, two Nintendo's Wiimotes, a pan-tilt camera head, and a computer for rendering wrapped images to the projector as shown in Figure 2. The configuration of projector and two Wiimotes are placed on the pan-tilt head that looks at infrared LEDs on the display surface. This configuration is designed to prove our algorithm which tracks four LEDs at the same time. To project an image onto a display surface, calibration between camera and projector for creating the virtual camera using image warping [8] needs to be done first. This virtual camera is created to match corresponding points between two Wiimotes' coordinates and convert those points into a projector's coordinate. Wiiuse library[9] is implemented for querying Wiimote's data in this prototype. Secondly, the targeted surface must be inside within the projector's frustum. To do that, it is very common to choose four corners of a display surface to be tracked. Figure 3 shows the locations of all four infrared LEDs in the display surface. The virtual camera then gives all corners' positions in 2D space to the system. OpenGL library is used to render textured graphics to fit the whole display surface.

Fig. 2. System overview of the proposed projected reality

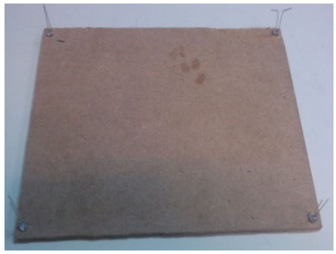


Fig. 3. Four diffusion infrared LEDs embedded on the corners of surface display

2.1 Nintendo's Wiimote

The Wiimote is one of the most common hacked game controllers developed by Nintendo Corporation. It includes many sensors such as accelerometer, infrared camera, and buttons. In this system, only the camera and Bluetooth connectivity are used.

The Wiimote can track up to the most brightest four infrared points at 100 Hz at 1024 × 768 pixels using multiobject tracking (MOT) engine [10]. The camera sensor has 45 degree horizontal FOV. In comparison with webcam at the same price, it gives more of 2.56 times resolution and 3.3 times refresh rate. Wiiuse library is used to retrieve camera data from the Wiimote via Bluetooth communication [9].

2.2 Pan-Tilt Head

The pan-tilt head is used for increasing workspace of perspective projection volume. In comparison with projector's FOV, it gives more of nine and four times on horizontal and vertical FOV, respectively. It consists of two digital servo networked motors (Robotis Dynamixel AX12+) which are connected to the computer via TTL to USB interface. Pan-tilt head is assembled in a simple direct drive form. To track the object, the origin of camera image on left top corner is changed to the center of image to find the offset which is determined as error on x and y axes. Centroid of four circles is also determined as delegated tracking point. The error in pixels is converted to an angular offset shown in [4] as following equations:

$$\Delta Pan = \tan^{-1}\left(\left(X_{pos} - \frac{X_{res}}{2}\right) \times \frac{\tan\frac{FOV_{hor}}{2}}{\frac{X_{res}}{2}}\right) \quad (1)$$

$$\Delta Tilt = \tan^{-1}\left(\left(Y_{pos} - \frac{Y_{res}}{2}\right) \times \frac{\tan\frac{FOV_{hor}}{2}}{\frac{Y_{res}}{2}}\right) \quad (2)$$

Where (X_{pos}, Y_{pos}) represents the centroid, X_{res} and Y_{res} are the size of image resolution. The angle for motors is obtained from the sum of the resultant values ΔPan and $\Delta Tilt$ and each present motor position. The absolute pan and tilt angles are sent to the motor controller when angular offset is more than a threshold to reduce unnecessary update. The speed of motor is varied between 2 rpm and 30 rpm accordingly to the motor's angular offset.

2.3 Virtual Camera

The simple model of location tracking system is shown in Figure 1. The real world object's position is at point O. It appears at point P on the camera plane. If the projector generates graphics at point P, the graphics is also displayed at point O. One of research's goals is to make a virtual camera with three constraints: i) this virtual camera is at the same as a projector's location and alignment, ii) horizontal and vertical FOVs are the same as projector's FOVs, iii) image plane has the same x and y axes as projector plane's axes. It can done by measuring object's position on x axis by using image obtained from camera which is translated on the vertical direction as shown in Figure 4. The previous process is also done for object's position on y axis in the same way. Two corresponding obtained points from two cameras are matched and used to determine a point in the virtual camera plane.

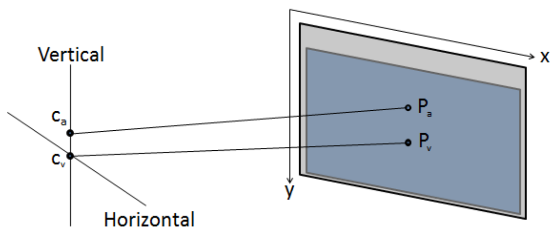


Fig. 4. C_v is position of the virtual camera. C_a is position of the actual camera. P_v and P_a are projections of the same object in the virtual and actual camera planes, respectively.

The first camera is placed closely to the projector with a little bit translation on the vertical axis as shown in Figure 4. The image from the vertical translated camera is warped and clipped to fit the projector's plane. Therefore, the projection points of the same object from the different camera planes, P_a and P_v, have the same coordinate in x axis for any depth. This process is also done for the second camera that is translated on horizontal axis. This gives coordinates in y axis on the virtual camera plane.

In the complete system with two Wiimotes as shown in Figure 5, P_a and P_b are points on the vertical camera and horizontal camera, respectively. (x_a, y_a) and (x_b, y_b) are coordinates of P_a and P_b, respectively. Equation 3 is used to calculate P_v which is the corresponding coordinate of the virtual camera plane.

$$P_v = (x_a, y_b) \tag{3}$$

Fig. 5. Two Wiimotes attached closely on both horizontal and vertical axes of portable projector

2.4 Image Matching

In [11], experimental results of fundamental matrix matching and square matching methods were compared. They took two sets of four points to determine the correspondence between the two images into the matched pair. The square matching method is simpler and has less complexity than the fundamental matrix matching which calculates twenty-four possible pairing for four points. Square matching method is also implemented in this paper. However, this algorithm is broken down when the points are rotated to the vicinity of the quadrant boundaries or rhombus quadrilateral. For carry this problem, all points need to be rotated a little bit around the centroid. The improvement of square matching method is implemented in this research to maintain its simplicity and can be adapted to any quadrilateral. First, the centroid, $C = (C_x, C_y)$, is calculated from a set of four points. If points are rhombus quadrilateral, they are rotated around the centroid by 5 degrees. Secondly, point id, $N = \{0,1,2,3\}$, is assigned for every point $P = (P_x, P_y)$ as shown in Equation 2.

$$N = u + v \tag{4}$$

$u = \{0,1\}$ is used to divide points into left or right sides of the centroid by a vertical line. If P_x is less than C_X, u is 0 otherwise u is 1. In the same way, $v = \{0,2\}$ is used to separate points into below and above sides of the centroid by a horizontal line. if P_y is less than C_y, v is 0 otherwise v is 2. The perpendicular of horizontal and vertical lines at the centroid divides each of the four IR points into unique quadrants by distinct point id. The same id of each image is corresponding point between images from two cameras. This algorithm has a limit to give up to accurate 4 points. As the matching points increase, fundamental matrix matching can be implemented to work with any number of points.

2.5 Coordinate Calibration between Two Wiimotes and a Projector

The virtual camera is the result of collection and calibration of two quadrilaterals from Wiimotes onto targeted quadrilateral of projector using projective mapping method[8]. It maps two Wiimote cameras' coordinate onto a projector's coordinate as shown in Figure 6. For example, if infrared LEDs appear in the projection frustum, the infrared LED positions are used to compute corresponding projector pixels projected on the positions of infrared LEDs. This method requires four triplets of corresponding points that are collected by four-point calibration process as shown in

Figure 6. It is a typical calibration process for any touch-screen system. First, four known locations are projected as the crosshairs on four corners of projector's plane. Secondly, an infrared LED is pointed at each of these crosshair's locations to obtain triplets of corresponding points. Finally, two warping matrices are computed by solving Gaussian elimination. In projective geometry, a 2D real point (x, y) is represented by the homogeneous vector $p = (x', y', w)$ where $(x, y) = (x'/w, y'/w)$ for $w \neq 0$. Given horizontal and vertical warping matrices, M_h and M_v, a vertical camera point $P_v = (s, t) = (s', t', 1)$, and a horizontal camera point $P_h = (u, v) = (u', v', 1)$. Hence, the projector point, P_0, can be obtained using Equations 3-5:

$$(x'_v, y'_v, w_v) = (s', t', 1)M_v \tag{5}$$

$$(x'_h, y'_h, w_h) = (u', v', 1)M_h \tag{6}$$

$$P_0 = (x_v, y_h) = (x_v, /w_v, y_h/w_v) \tag{7}$$

For a projector's coordinate in Equation 5, x and y values are obtained from the vertical and horizontal cameras, respectively.

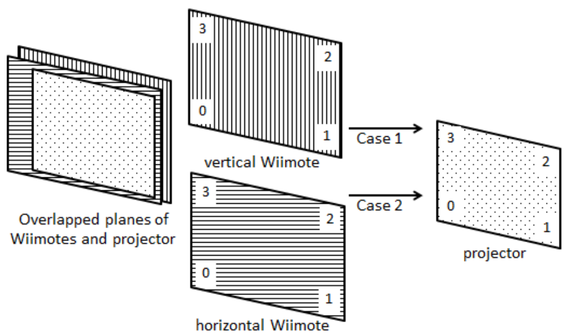


Fig. 6. Overlapped planes from horizontal Wiimote, vertical Wiimote, and projector plane

2.6 User Interface

The user interface (UI) is designed for proofing the concept of the virtual camera. The UI is implemented using OpenGL library. It has two main windows which are a full screen graphics window for projector and a command line window for receiving user's requests. Before using the system, user presses the "C" key to calibrate Wiimotes and a projector using four-point calibration method. User then chooses augmented data such as static image and video as shown in Figure 7. When the display surface is activated in the workspace, the full screen graphics image will be displayed to provide augmented data selected by user's commands. If the object is moved far away from the center of image, the pan-tilt head will move to track that object.

Fig. 7. From left to right: video image projected on moving display. Address information augmented onto a post box.

3 Experimental Results

Active location tracking system is implemented using a portable projector (3M MP150), two infrared cameras (Wiimotes), and a pan-tilt head (Robotics Dynamixel AX12+). The images are at a resolution of 640 by 480 pixels. Four 940 nm diffusion infrared LEDs are attached to four corners of the display surface. All system components are interfaced with a Microsoft's Windows 7 laptop with an Intel Core i5 M460 at 2.53GHz. The location tracking method is evaluated in three areas: i) Geometric Proof ii) Projection Accuracy iii) Processing throughput.

3.1 Geometric Proof

Figure 8 is drawn using Thai Geometer's Sketchpad. It represents boundary of projection volume, object in real world space, object position in image plane separately for each axis. If camera is translated on the vertical axis, only image on y axis has an error as shown in the right side of Figure 8. In this condition, images between projector and camera are parallax. The intersection region can use camera for measuring value in x axis as shown in the left side of Figure 8. In the other word, if camera is translated on the horizontal axis, it can measure the location in y axis.

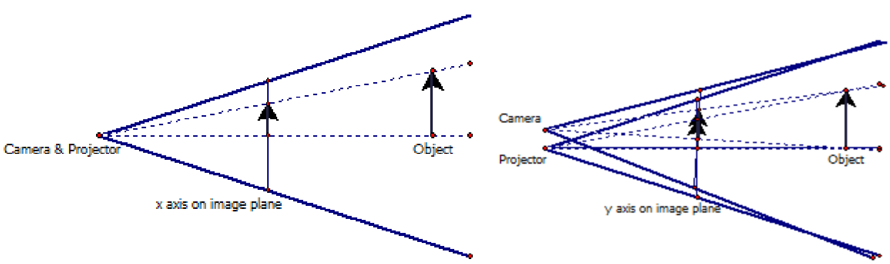


Fig. 8. Camera's translation on vertical axis. Left: top view shows the image on x axis projected on the same position as camera's. Right: left view shows the image on y axis overlapping between projector and camera.

3.2 Projection Accuracy

Projection accuracy is critical for projected reality. In experiments, the accuracy is evaluated by the error from the location tracking using projection crosshair pattern onto the target in Figure 9. The target consists of multiple radius circles with shared center which is infrared LED. There are 10 circles. The first one has a radius of an infrared LED's which is 0.25 cm. The range of each circle interval in target is 0.25 cm. The projection volume is separated into a quadrant which has the origin in the center of image. The test point is assigned randomly for 10 points per quadrant and is measured the error using the circle interval that shows the center of the crosshair. This experiment was run three times with different distances from projector to the plane at 60, 80, and 100 cm. From experimental results, most of projected points are close to the center of the target and placed in the first circle interval shown in Table 1. The class interval arithmetic mean is 0.25 cm.

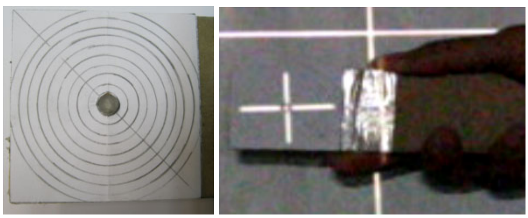


Fig. 9. Left: diffusion LED is pitched at the center of class interval. Right: the error is measured by projected crosshair.

Table 1. Projection Error

Error Intervals (cm)	Frequency(f)
0.00-0.25	99
0.25-0.50	21
0.50-1.00	0

3.3 Processing Throughput

The expected throughput is more than the typical projector refresh rate at 60 fps. This is important for real-time projected reality with moving object. Two sections affect the processing throughput which are input and output section. The effect of input section depends on the refresh rate of the camera and the latency of matching. The effect of output section depends on the latency of warping and rendering graphics in OpenGL. In first section, the camera refresh rate is fixed at 97 fps. The matching algorithm is used only when the new infrared is activated or disappeared. In the other word, it is typically executed only when the surface display comes into the workspace for the first time. Hence, this overhead does not change the camera throughput significantly. For the output section, Table 2 shows the number of frames processed per second when it renders different kinds of graphics. The throughput decreases as the complexity of graphics texture increases. However, the throughput is still higher than the projector's refresh rate. This proposes system has ability to do a real time location tracking.

Table 2. Latency of texture rendering

Texture Type	Frame Rate (fps)
4 crosshair (non-texture)	196
Image	147
Video	123

4 Conclusions and Future Work

This system can augment graphics image onto the moving object using two Wiimotes and a portable projector with a pan-tilt head. Two Wiimotes are placed a little bit of translation on horizontal and vertical axes to track LEDs placed in each corner of the display surface. The system obtains 2D position from two Wiimotes, matches corresponding points, computes the 2D position of point in the projector's plane, and displays texture accurately fit to the surface. This research shows the use of two camera calibration algorithm to directly map 2D camera's coordinate into projector's coordinate using perspective projection geometry property. The camera inside the Wiimote has a limit of tracking capability for only four LEDs which is insufficient to track the complex surface. This would be an interesting avenue for future work to deal with complex surface tracking.

Acknowledgments. This research work is financially supported by the National Science and Technology Development Agency, Thailand.

References

1. Pinhanez, C.: The everywhere displays projector: A device to create ubiquitous graphical interfaces. LNCS, pp. 315–331 (2001)
2. Bandyopadhyay, D., Raskar, R., Fuchs, H.: Dynamic shader lamps: Painting on real objects. Citeseer (2001)
3. Karitsuka, T., Sato, K.: A wearable mixed reality with an on-board projector, p. 321. IEEE Computer Society, Los Alamitos (2003)
4. Ehnes, J., Hirota, K., Hirose, M.: Projected augmentation-augmented reality using rotatable video projectors, p. 35. IEEE Computer Society, Los Alamitos (2004)
5. Park, H., Lee, M.H., Kim, S.J., Park, J.I.: Surface-independent direct-projected augmented reality. In: Computer Vision–ACCV 2006, pp. 892–901 (2006)
6. Lee, J.C., Dietz, P.H., Maynes-Aminzade, D., Raskar, R., Hudson, S.E.: Automatic projector calibration with embedded light sensors, pp. 123–126. ACM, New York (2004)
7. Lee, J., Hudson, S., Dietz, P.: Hybrid infrared and visible light projection for location tracking, pp. 57–60. ACM, New York (2007)
8. Heckbert, P.: Projective Mappings for Image Warping. Image-Based Modeling and Rendering 15-869
9. Laforest, M.: Wiiuse API Overview (2008)
10. Lee, J.C.: Hacking the nintendo wii remote. IEEE Pervasive Computing 7, 39–45 (2008)
11. Wang, D., Huang, D.: Low-Cost Motion Capturing Using Nintendo Wii Remote Controllers

Fast Prototyping of Virtual Replica of Real Products

Francesco Ferrise and Monica Bordegoni

Politecnico di Milano, Dipartimento di Meccanica,
Via La Masa 1, Milano, Italy
{francesco.ferrise,monica.bordegoni}@polimi.it

Abstract. The ability to capture customers' needs and the voice of customers, and to translate them into a set of product specifications that at best satisfy the target customers has increasingly become a key element of business strategy. The common practice consists in evaluating products at the end of the design process through physical prototypes with the participation of users and potential customers. The same practice can be implemented by using virtual replica of real products, reducing cost and time necessary to build some variants. The paper presents a methodology for the development of the virtual prototype of a piece of furniture, produced by a company that is interested in studying how customers perceive and evaluate some variants of the hinge mechanism. The virtual prototype has been implemented using a tool for virtual reality applications oriented to non-expert programmers. The modularity and flexibility of the approach used for implementing the virtual replica has allowed us to re-use the components, and to easily change the parameters, also during the test activities.

Keywords: Virtual Products, Virtual Prototyping, Fast Prototyping.

1 Introduction

One of the recent trends in interactive product design [1] is the study of the perceptual aspects of the products, meant as the way to drive the subjective impression that customers have when interacting with the product. While sometimes in the industrial design practice the perception is associated with pure visual elements such as shapes, colors, and textures, actually also sound and haptic cues concur in creating the overall perceptual feedback elicited by the product. This happens especially for those products where interaction is not based on pure visualization, but even on touch and auditory aspects. When observing customers in front of products of different brands with similar prices and technological contents, it can be noticed that they begin to interact with them in order to evaluate the product on the basis of the emotions provoked and felt. It happens for different kinds of commercial products, from textiles to household appliances, to cars. This is what is sometimes recognized as "first contact" with the product that is important since it influences the potential customer at the moment of purchase, together with the price and other more objective aspects [2].

Designing the perceptual aspects of products can be a successful strategy when no more technological contents can be added to a product to increase the objective value. Sometimes the perceptual aspects of the products, and in particular those connected to

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 318–326, 2011.

the sense of touch can be evaluated only when a physical prototype that is much as possible similar to the final product is available, and sometimes it happens too late in the product development process to perform some significant changes with a small effort in terms of money.

In years Virtual Prototyping, meant as the substitution of physical prototypes with their virtual replica (Virtual Prototypes, VP) in some testing activities of the product development process [3], has become a diffused practice in the industrial context, in order to reduce costs and time required to test different variants of the same product. For years the fast evolution of visualization technologies, both from the hardware and software points of view with respect to the haptic and sound ones has constrained the testing activities concerning the pure visualization aspects. In the recent years haptic and sound technologies have gained a technological level so they can be used to enlarge the number of testing activities that can be performed in a VR context to those including the senses of touch and hearing, and so even to simulate a perceptual feedback much more complex and similar to the real one.

In the paper we describe a case study where we have defined a methodology for the development of a fast virtual replica of a piece of furniture, specifically a cupboard. The case study has been proposed by a company interested in studying the hinge mechanism, by testing how users perceive various variants of the mechanism. Therefore, we have been asked to build a virtual prototype of the cupboard, simulating the functional properties of the hinge system, including the possibilities of rapidly switching between alternatives, or changing the physical behavior and responses. The virtual prototype consists of a realistic stereoscopic visualization of various kinds of cupboards, integrated with the possibility of opening and closing the doors through the use of a general-purpose 6DOF haptic device. The user can see a stereo view of the various cupboards, and can open and close their doors, by feeling forces returned during the interaction.

The paper is structured as follows: Section 2 describes the main achievements in similar research activities and highlights the main novelties of the work presented in this paper, Section 3 describes the case study, Section 4 reports the main steps of the implementation of the fast virtual replica of the product, and finally Section 5 discusses of the results obtained.

2 Related Works

The involvement of final users in the product development process is particularly important. In fact, a product might be the best from the technological point of view, but may elicit a very poor emotional response in the users. Therefore, the study and understanding of the emotional response and appreciation of the product might concur in designing better products in terms of marketing success [2, 4, 5]. One of the trends that are spreading in the product design process is substituting physical prototypes that are necessary to perform tests and evaluations with their virtual replica [3].

Since visualization technologies have been largely studied and refined in the years, more than the haptic ones, most of the tests performed have been limited to visualization aspects. In [6-8] are described some examples of tasks concerning product evaluation that can be performed when visualization is the unique available communication channel. These virtual prototypes consist of realistic, sometimes high fidelity and real-time, visualization of the products. Advanced stereoscopic visualization technologies allow for a realistic visual perception of the virtual models. These environments are appropriate for the evaluation of aesthetic aspects of products, as variants of colors, textures and materials in different lighting conditions, but not for the evaluation of other features, especially those related to users' interaction. The possibility of physically touching and manipulating, and also operating and interacting with an object, is an important issue in the assessment of several consumer products. In fact, the evaluation of the physical interaction by means of “non-physical” components is definitely poor and does not provide useful results.

Some simple haptic devices, such as knobs, have been used for testing user interaction with the interface components of consumer products [9, 10]. The user is able to evaluate different pre-defined haptic behaviors of the knob and choose the favorite one. Another example concerns the use of haptic devices and physics-based simulations to reproduce human interaction with textiles [11]. Here the VR environment is purely haptic-based and does not include any visualization. In [12] we present a multimodal system based on haptic, sound and visualization technologies to test human interaction with the Virtual Prototype of a commercial washing machine. Haptic feedback is returned through a 6DOF haptic device equipped with a generic handle, and some recorded sounds are played when collisions occur. This work has inspired the case study described in this paper.

3 Description of the Case Study

The case study has been provided by the SALICE company (www.salice.com), which design and produce that is interested in studying the hinge mechanism of cupboards doors, by testing how users perceive various variants of it. Therefore, we have been asked to build a virtual prototype of the cupboard, simulating the functional properties of the hinge system, including the possibilities of rapidly switching between alternatives, or changing the physical behavior and responses.

In particular we were interested in simulating the behavior of three different hinge systems: the first one (Figure 1-a) is the folding door system that consists of a double door in order to provide access to a double cupboard. A composed movement permits the user to open the double cupboard only by interacting with the bottom door. The second one is the flap door system with push opening (Figure 1-b) that consists of a single flap door provided with a push opening system that allows the user to open the cupboard by simply pushing. The third one (Figure 1-c) is the flap door system with normal opening. The user must pull until a certain opening point and then the lifting system opens the door. To close the door the user must push up to the same point where the torque returned by the lifting system changes, and the door gets closed.

(a) (b) (c)

Fig. 1. The three cupboards used as case study: a) the folding door, b) the flap door with push opening, c) the flap door with normal opening

The manufacturers provided us both the real prototype of the three hinge systems and the CAD models. CAD models are useful to create the visual prototype and to study the hinge mechanisms dynamic behaviors, in order to reproduce them into the virtual environment. The real objects are useful in order to adapt the perceptual experience to what happens in the real product, and so to adapt forces and to match the real ones.

4 Methodology to Create a Fast Virtual Replica

In this Section we describe the methodology that we have used to implement a fast virtual replica of the real product. First we discuss the choice of the hardware and software equipment, and then we describe how to simplify the implementation, while creating an overall perceptual experience that is as much as similar to the real one.

4.1 Choice of the Hardware and Software Equipment

Since the aim of the application is to allow a non-expert users to create rapidly a virtual prototype to interact with, it is obvious that the first step is to choose commercial tools, both from hardware and software point of view. Virtual Reality offers a wide variety of hardware and software technologies that have specifically been developed to communicate through three sensorial channels: touch, vision and hearing [13-15].

Since flexibility is one of the key points of our application we base our implementation on a VR setup consisting of general-purpose devices. Specifically the development of the application is based on the following hardware and software commercial tools (Figure 2):

- the *Haption Virtuose* 6DOF haptic device (www.haption.com). Despite being limited in terms of magnitude of forces and torques that it can render, this haptic device is provided with a big working volume if compared with the other haptic devices available on the market, and returns both forces and torques;

- a retro-projected wall display *Cyviz*, that is based on linear polarizers mounted on two projectors and on some lightweight glasses worn by the user, for the stereoscopic rendering of the scale model of the three cupboards (www.cyviz.com). We have preferred a wall display to several commercial available HMDs since it is the less intrusive and can be used for a longer period [16];
- three infrared optical *AR-Tracking* cameras for the detection of the user's point of view (www.ar-tracking.de). In this way the visual exploration of the object becomes more natural;
- the *3DVIA VirTools*, which is a VR development environment (www.3dvia.com).

We have decised to use the VirTools development environment first because it is well integrated with the 3DVIA CAD tool that has been used for implementing the digital models, that is SolidWorks. VirTools is based on a simple building-block programming paradigm and is well suited to be used even by non-expert programmers. This is important since the aim of the work described in the paper is to create a methodology to build a fast virtual replica of a real product, in order to perform some qualitative analysis on users' perception of it. This fast prototype can be used even in the early stage of the product development process where some detailed decisions have not been taken yet. At this stage since several different variants of the same product could be necessary, what is important is to be flexible enough to switch from a behavior to a completely different one without wasting too much time.

Fig. 2. The hardware setup used to implement the case study consisting in: an optical tracking system, a retro-projected wall display and a 6DOF haptic device equipped with a generic handle

4.2 Simplified Visuo-Haptic Interaction

VirTools is well integrated with the hardware equipment that we have selected to implement our case study. VirTools has some limits in terms of haptic control algorithms if compared to other specific open source haptic libraries such as CHAI3D (www.chai3d.org) and H3DAPI (www.h3dapi.org). Anyway it allows a fast creation of a visuo-haptic interaction environment.

The visuo-haptic environment should allow the user to naturally interact with the virtual prototype. It should include a scale model of the object that is necessary since using wrongly displayed visual information can alter the overall perceptual feedback. Then it should not contain any GUI elements if not necessary that might distract a potential user while performing some tests on the prototype. The interaction should be as much as possible intuitive so as not to require a long familiarization with the environment. The VR environment should be transparent as much as possible and this is important since potential testers can be common users, who do not necessarily have confidence with VR technology.

Our visuo-haptic environment works in the following way: the user holds the handle of the haptic device and moves a visual avatar of his hand that has been reproduced in the virtual environment. The haptic device is used as input device until the hand reaches one of the snapping areas illustrated in Figure 3. When the hand is inside one of the snapping areas, it automatically changes its shape by mimicking a real hand and is automatically attached to the door. The user is now able to open and close the doors, and the haptic device returns forces and torques to the user's hand. The avatar of the hand has been introduced so as to make interaction more realistic and even to constrain the user to be visually guided during haptic interaction [17].

A combination of forces and torques is returned to the user's hand depending on which cupboard she is interacting with. For example in the case of the folding door (Figure 1-a) and of the flap door with normal opening (Figure 1-c) a simple combination of a constant force and a constant torque is returned to the hands of the user. The direction and orientation of the forces has been computed from the CAD models and then imported in VirTools, while the magnitude has been roughly computed from simplified mechanical laws and then finely tuned to be as much as possible similar to the real cupboards. In the flap door with push mechanism a localized effect is added when the door reaches the closing angle. In the case of the push mechanism (Figure 1-b) it has been implemented as a two states system: one where the system maintains the door closed and the other one when allows the lifting system to open the door. This state changes every time the user pushes the door up to the maximum closing angle.

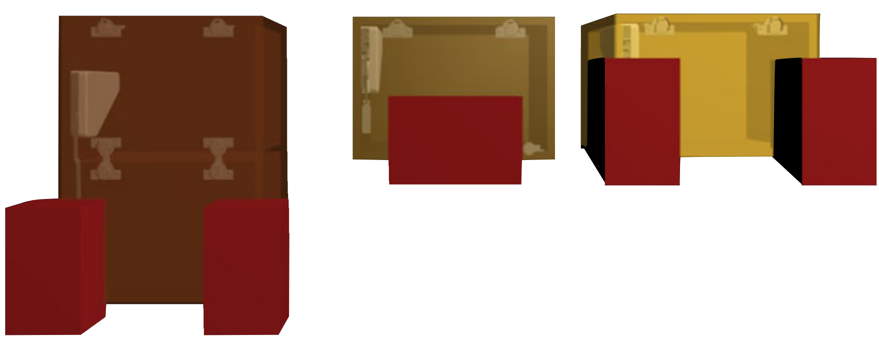


Fig. 3. Some haptic snapping effects have been introduced to simplify interaction with the Virtual Prototype. When the virtual hand is inside one of the red blocks it automatically changes its shape mimicking a real hand interacting with the same object and is attached to the door returning to the user forces and torques.

The force model implemented in VirTools, despite being very simple, has been implemented in a way that it can be easily modified upon user's requests. In this way during the design review with potential customers some coarse mistakes can be rapidly corrected, by adapting the force feedback to user preferences and printing out the new force and torque values that become design inputs as described in [18].

In order to reproduce the rotational limits of the hinge mechanism some small blocks have been introduced in the virtual environment. In this way the limits are obtained by the collision of the door that is rotating and the block that is fixed (see Figure 4). These blocks are not visually rendered.

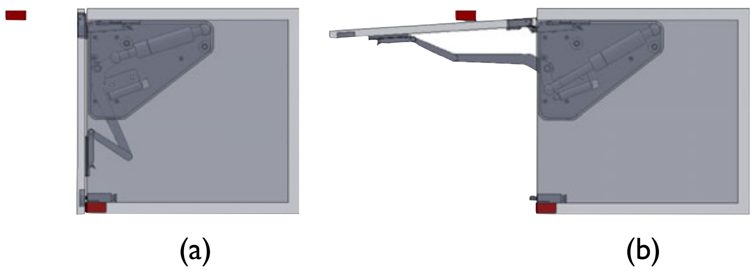


Fig. 4. The rotational limits of the hinge mechanism have been simply reproduced by introducing some collisions between the unique door that is physicalized and some fixed small blocks that are not visually rendered.

Haptic feedback is not computed on the basis of the visual model through some complex simulation algorithms, as it usually happens in physics-based simulations. It is indeed completely independent. For instance, in the three examples the only element that has some physical properties of mass and surface stiffness is the door, which is grasped by the avatar of the hand. It is used to simulate the rotational limits by colliding with the limit blocks (Figure 4). All the other elements, as for example those that compose the hinge mechanism, are purely visual and are constrained to follow the door that is rotating, with the same movements of the real object.

In the development of the visuo-haptic interaction environment we concentrated mainly on the reproduction of an overall perceptual experience as much as possible similar to the one the user experiences in the real world. The perceptual experience in the VR environment is the combination of a natural visual exploration and the haptic feedback. The visual exploration of the prototype is granted by the use of the scale model reproduced in stereoscopy, and the tracking of the user's point of view is allowed by the optical tracking system. Then the interaction is made more natural thanks to the use of the haptic device.

In the case study described in the paper we have used a real object to tune the haptic parameters, since our initial aim was to create a virtual replica of the real cupboards. Conversely, if our aim is to design haptic interaction we can reverse the procedure and design the interaction without using the real prototype as described in [18].

5 Discussion and Conclusion

The paper has presented a methodology to create a fast virtual replica of a real product. A fast prototype is important when some preliminary testing activities should be run early in the product development process, when several strict decisions have not yet been taken. In particular our virtual prototype can be used to perform some testing activities concerning human interaction, concerning user's perception of the product. It is based on a simplified visuo-haptic interaction environment. A haptic device returns to the user's hand forces and torques that can be easily extracted from the CAD model, and the user is free to visually explore the object. He can ask some modifications of the haptic behavior, and these values can be used as design specifications for a new product. We made use of general-purpose commercial software and hardware tools since the aim of the work was to demonstrate how even non expert users could be able to create their own VR interaction environment. But it is clear that spending much effort on customizing the devices for each sense we can make the environment usable even for a more robust and quantitative analysis. Anyway the application at this stage can be used to correct some coarse mistakes since the very beginning of the product development process, since it allows operators to switch among different design variants in short time.

Acknowledgments. The authors would like to thank Joan Segarra Gasull for his precious contribution during the development of the case study and the SALICE Company for providing us the products used as case study.

References

1. Ulrich, K., Eppinger, S.: Product Design and Development. McGraw-Hill, New York (2008)
2. Zeithaml, V.A.: Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. Journal of Marketing 52(3), 2–22 (1988)
3. Wang, G.G.: Definition and review of virtual prototyping. Journal of Computing and Information Science in Engineering 2(3), 232–236 (2002)
4. Norman, D.A., Draper, S.W. (eds.): User Centered System Design: New Perspectives on Human-computer Interaction, 1st edn. CRC Press, Boca Raton (1986)
5. Hvam, L., Mortensen, N.H., Riis, J.: Product Customization. Springer, Heidelberg (2008)
6. Verlinden, J., Horvath, I., Nam, T.-J.: Recording augmented reality experiences to capture design reviews. International Journal on Interactive Design and Manufacturing 3(3), 189–200 (2009)
7. Santos, P., Stork, A., Gierlinger, T., Pagani, A., Paloc, C., Barandarian, I., Conti, G., Amicis, R.D., Witzel, M., Machui, O., Jimenez, J., Araujo, B., Jorge, J., Bodammer, G.: Improve: An innovative application for collaborative mobile mixed reality design review. International Journal on Interactive Design and Manufacturing 1(2), 115–126 (2007)
8. Schmalstieg, D., Fuhrmann, A., Hesina, G., Szalavari, Z., Encarnacao, L.M., Gervautz, M., Purgathofer, W.: The studierstube augmented reality project. Presence: Teleoperators and Virtual Environments 11(1), 33–54 (2002)
9. Bordegoni, M., Colombo, G., Formentini, L.: Haptic technologies for the conceptual and validation phases of product design. Computers & Graphics 30(3), 377–390 (2006)

10. Kim, L., Han, M., Shin, S. K., Park, S. H.: A haptic dial system for multimodal prototyping. In: 18th International Conference on Artificial Reality and Telexistence – ICAT (2008)
11. Magnenat-Thalmann, N., Volino, P., Bonanni, U., Summers, I.R., Bergamasco, M., Salsedo, F., et al.: From physics-based simulation to the touching of textiles: The haptex project. The International Journal of Virtual Reality 6(3), 35–44 (2007)
12. Ferrise, F., Bordegoni, M., Lizaranzu, J.: Product Design Review Application Based on a Vision-Sound-Haptic Interface. In: Nordahl, R., Serafin, S., Fontana, F., Brewster, S. (eds.) HAID 2010. LNCS, vol. 6306, pp. 169–178. Springer, Heidelberg (2010)
13. Burdea, G., Coiffet, P.: Virtual Reality Technology, 2nd edn. John Wiley, New York (2003)
14. Grünwald, M. (ed.): Human Haptic Perception - Basics and Applications. Birkhäuser, Switzerland (2008)
15. Farnell, A.: Designing Sound. The MIT Press, Cambridge (2010)
16. Sharples, S., Cobb, S., Moody, A., Wilson, J.R.: Virtual reality induced symptoms and effects (vrise): Comparison of head mounted display (hmd), desktop and projection display systems. Displays 29(2), 58–69 (2008); Health and Safety Aspects of Visual Displays
17. Plaisier, M.A., Kappers, A.M.L., Tiest, W.M.B., Ernst, M.O.: Visually Guided Haptic Search. IEEE Transactions on Haptics 3(1), 63–72 (2010)
18. Bordegoni, M., Ferrise, F., Lizaranzu, J.: Use of Interactive Virtual Prototypes to define product design specifications: a pilot study on consumer products. In: Proceedings of IEEE – 2011 International Symposium on VR Innovation – ISVRI (2011)

Effectiveness of a Tactile Display for Providing Orientation Information of 3d-patterned Surfaces

Nadia Garcia-Hernandez, Ioannis Sarakoglou, Nikos Tsagarakis , and Darwin Caldwell

Italian Institute of Technology, Advanced Robotics Department, Via Morego 30, 16163 Genova, Italy
{nadia.garcia,ioannis.sarakoglou, nikos.tsagarakis,darwin.caldwell}@iit.it

Abstract. This paper studies the effectiveness of a tactile display in providing information about the orientation of 3d-patterned surfaces. In particular, it investigates the perception of the orientation of sinusoidal gratings rendered through the display in a passive guided touch modality. The results of this study have revealed that participants could successfully perceive variations in the orientation of the rendered sinusoidal gratings. Moreover, they indicate a small difference in the perception of orientation between touching virtual gratings and touching real gratings.

1 Introduction

In recent years, there has been increasing interest in the integration of tactile devices in virtual reality and teleoperated systems [1], [2]. This integration would allow feeling objects' tactile properties, such as local shape [3], texture [4], and temperature [5]. However, due to limitations of the current actuation technologies, the existing tactile devices are unable to provide all the tactile sensations that humans experience when touching real objects. The way to find out which tactile sensations are able to be simulated by a tactile display, and to quantify how close they are to real touch under natural conditions, is to conduct studies that quantify users' tactile perception ability.

This work aims to investigate how accurate the tactile display presented in [6] provides information about the orientation of 3d-patterned surfaces, in particular surfaces with a sinusoidal profile. In touch, tactile orientation cues provide critical information about how stimuli are positioned on the fingers which is important for grasping and lifting objects. It also provides important information about tactile shape. Psychophysical studies have shown that humans have a high capacity to discriminate the orientation of shapes and gratings indented into the finger pad [7]. The display used in this work is compact, lightweight and consists of 4x4 vertically moving pins remotely actuated by miniature DC motors (Fig. 1). To provide an objective reference for the evaluation of the rendering and perception of oriented gratings through the tactile display, it was also investigated the perception of orientation of real gratings (explored using the bare finger). The results presented in this paper reflect the effectiveness of the tactile display to provide orientation information of small-scale shapes.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 327–332, 2011.

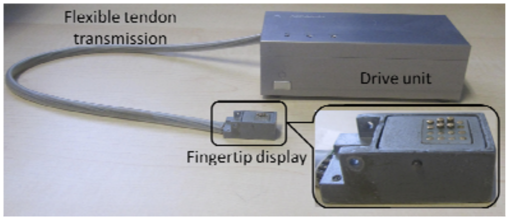


Fig. 1. Fingertip tactile display of 4x4 vertically moving pins

2 Materials

2.1 Tactile Display and Experimental Setup

The tactile display shown in Fig. 1 consists of 16 vertically moving tactors (4x4), it has a tactor spacing of 2mm (centre to centre) and a tactor diameter of 1.5mm. The tactors are spring loaded by 0.61N/mm springs and have a nominal amplitude range of 0-2mm. The available force is in the range of 0.61N-1.83N between maximum and minimum displacements. Control of the tactors is continuous over their motion range in a position control fashion. The remotely located actuators and the flexible tendon transmission allow the display component to have a small size, which makes it easy to attach it to the fingertip. Moreover, it allows the user to freely explore virtual surfaces in a natural way utilising a large part of the finger' workspace.

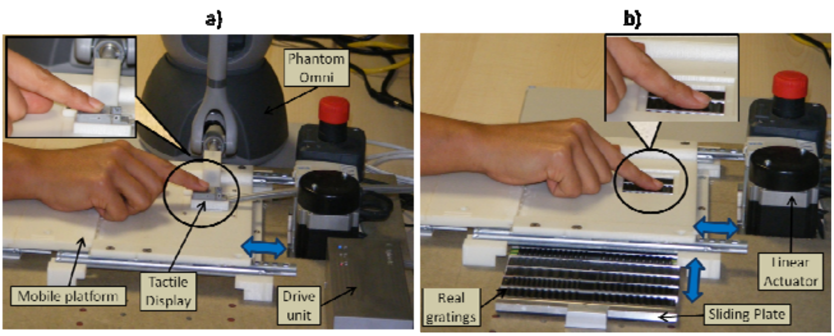


Fig. 2. a) Setups for the tactile exploration of a) virtual gratings and b) real gratings

The passive-guided exploration of the virtual and real gratings was performed through the setups shown in Fig. 2. During the tactile exploration, the participants lied their arm-hand down on a mobile platform and placed their index finger either on top of the tactile display for the exploration of virtual gratings (Fig. 2a) or on top of the real gratings (Fig.2b). The tactile exploration was performed by controlling the velocity and linear displacement of the mobile platform. This was achieved by using a SMC linear rodless actuator (Series E-MY2B) which allows 10 different velocities (10, 20, 30, 40, 50, 75, 100, 300, 500 and 1000mm/sec).

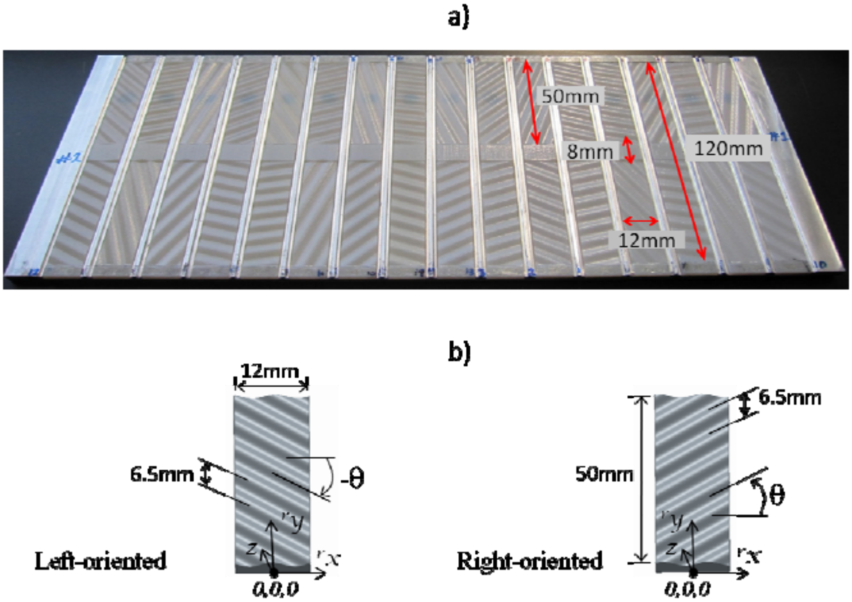


Fig. 3. a) Plastic strips of sinusoidal gratings and b) Coordinates for the tactile rendering of the oriented sinusoidal gratings

2.2 Virtual and Real Gratings

Both the virtual and the real gratings occupied an exploration area of 12mm wide by 50mm long and had a wavelength of 6.5mm and amplitude of 0.6mm. The real gratings were printed on plastic strips using a high-resolution 3D printer. Virtual gratings were displayed by controlling the vertical displacement of the tactile display pins according to the calculated shape profile of the simulated gratings.

3 Methods

3.1 Subjects

Seven right-handed subjects participated in this study. Each of them performed the test for the discrimination of the virtual and real gratings using two velocities of exploration 20mm/sec and 40mm/sec. All subjects chose their index finger to perform the test and they were not allowed to look at the gratings while performing the tactile exploration.

3.2 Gratings Rendering

The virtual gratings oriented at a tilt of θ were recreated along the ${}^{r}y$-axis, as it is shown in Fig. 3b. The origin of the ${}^{r}y$-axis is the start point of the tactile exploration and corresponds to the origin of the surface. As the subjects' finger (on top of the tactile display) moved along the ${}^{r}y$-axis, the position and displacement of the pins was calculated. The position of each pin n (${}^{r}P_n({}^{r}x_n, {}^{r}y_n)$), was calculated first by rotating the pin array θ degrees (the array central point C(*0,y*) is rotated around *z*-axis) and then by calculating their distance with respect C(*0,y*), as is shown in Eq. (1).

$$ {}^{r}P_{n}\left({}^{r}x_{n},{}^{r}y_{n}\right)=P_{n}(x_{n},y_{n})\cdot Rot(z,\theta)\,. \tag{1}$$

The height of each pin n (h_n) is calculated based on its position (${}^{r}P_n({}^{r}x_n,{}^{r}y_n)$), as it is shown in the next equation.

$$ h_{n}\left({}^{r}P_{n}\right)=K\left[\frac{1}{2}+\frac{1}{2}\cos\left(\frac{2\pi}{\lambda}{}^{r}P_{n}-\pi\right)\right],\quad if\quad 0\leq{}^{r}P_{n}({}^{r}x_{n},{}^{r}y_{n})\leq 50;\;\cdot \tag{2}$$

where λ represents the spatial period of the corrugated surface and *K* represents the maximum amplitude of the surface (0.6mm).

3.3 Design and Psychophysical Procedure

Before starting a trial with the real gratings, the experimenter slid the sliding plate until the desired gratings were placed under the subjects' fingertip.

During a trial, participants were allowed to explore the virtual gratings (finger on top of the tactile display, Fig. 2a) and the real gratings (finger on top of the gratings, Fig. 2b) in one direction at the speed of 20mm/sec (for 5sec) and 40mm/sec (for 3sec). The gratings were left or right-oriented, see Fig. 3b, at one out of eighteen tilt angles ranged between -75° and 75°. At the end of each tactile exploration, participants replied verbally if the gratings were right or left-oriented.

In each exploration condition prior to the experimental trials a few practice trials were performed in order to let subjects become acquainted with the task. For the experimental trials, there were 10 repetitions for each tilt angle and velocity of exploration. This made a total of 320 trials for the virtual gratings (160 trials per exploratory speed) and 320 trials for the real gratings (160 trials per exploratory speed). All trials were randomized and accomplished by groups of 80 trials in different sessions. Subjects' performance was evaluated from the percentage of correct responses (PCR).

4 Results

Fig. 4 shows the mean PCR for each tilt angle using real and virtual gratings. Results exploring at 20mm/sec and 40mm/sec are shown in Fig. 4a and Fig. 4b, respectively. Fig. 4c summarises the results obtained at both velocities. Paired t-tests indicated that the mean PCR obtained for virtual gratings when exploring at 20mm/sec ($t(16)=3.35$, $p=0.004$) and at 40mm/sec ($t(16)=2.86$, $p=0.012$) differs significantly from the mean PCR obtained for real gratings. They also indicated that the mean PCR obtained for real ($t(16)=0.73$, $p=0.48$) and virtual gratings ($t(16)=0.44$, $p=0.67$) was not affected by the used exploratory velocity (20 and 40mm/sec).

The mean PCR decrease considerably and standard deviation increases for the tilt angles between -30° and 30° and for 60° and 75°. The decrease is greater in virtual gratings (from 98% to 52%) than in real gratings (from 98% to 67%).

The overall mean PCR across the exploration velocities showed that the perception of orientation of real gratings (89%) was more accurate than for virtual gratings (81%). Moreover, it was observed that the scores for left-oriented gratings were slightly greater than for right-oriented gratings, see the percentage values in bold in Table 1.

Table 1. Confusion matrix of percentage of correct answers for real and virtual gratings

		Exploration at 20mm/sec		Exploration at 40mm/sec	
		Right	Left	Right	Left
Answers for Virtual Gratings	"Right"	**78.1%**	21.9%	**78.6%**	21.4%
	"Left"	16.7%	**83.3%**	16.4%	**83.6%**
Answers for Real Gratings	"Right"	**89.5%**	10.5%	**87.3%**	12.7%
	"Left"	10.5%	**89.5%**	9.9%	**90.1%**

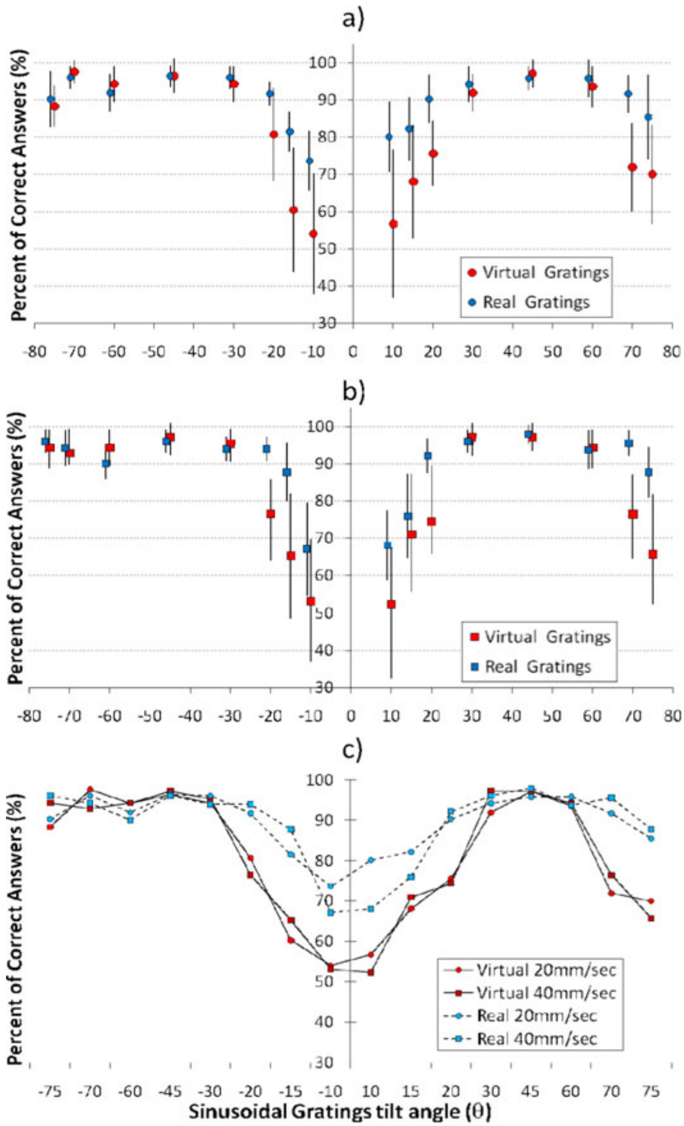


Fig. 4. Mean percentage of correct responses for each tilt angle using real and virtual gratings. a) Exploring at 20mm/sec, b) 40mm/sec and both 20 and 40mm/sec.

5 Discussion and Conclusions

The results of this study have revealed that when using the tactile display, participants could successfully perceive variations in the orientation of small-scale sinusoidal gratings. The discrimination performance of the participants was not affected by the tactile exploratory velocities used in this study. Moreover, it was found a small difference in the perception of orientation between virtual and real gratings. The small size of this difference is a very encouraging and somewhat surprising result. Given that several factors differ between touching a surface directly with the finger and through an array of pins, one might have expected a larger difference in discrimination performance. In particular, when touching real gratings a complete set of continuous forces (normal, shear and sliding forces) are present and are sensed by the users' fingertip. These force components should be expected to contribute to the overall tactile perception. However, when touching virtual gratings, only discrete normal forces (delivered by the tactile displays' tactors) are sensed by the users' fingertip. Secondly, the tactile display has a considerably smaller tactor density (25units/cm^2) in comparison to the human SAI mechanoreceptor density (70units/cm^2) that it mainly stimulated.

Finally, the presented data clearly suggest the promising use of this tactile display as an effective means for simulating virtual surfaces containing information of shape and orientation. Future work will involve extensive perceptual studies in order to provide a more complete perceptual evaluation of the tactile display.

References

1. Ottermo, M.V., et al.: Palpation instrument for augmented minimally invasive surgery. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 4, pp. 181–184 (2004)
2. Kontarinis, D.A., Howe, R.D.: Tactile Display of Vibratory Information in Teleoperation and Virtual Environments. Presence: Teleoperators and Virtual Environments 4, 387–402 (1996)
3. Kammermeier, P., et al.: Dynamic display of distributed tactile shape information by a prototypical actuator array. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 2, pp. 1119–1124 (2000)
4. Ikei, Y., et al.: Texture presentation by vibratory tactile display-image based presentation of a tactile texture. In: IEEE Virtual Reality Annual International Symposium, pp. 199–205 (1997)
5. Yang, G.-H., et al.: Use of Simulated Thermal Cues for Material Discrimination and Identification with a Multi-Fingered Display. Presence: Teleoperators and Virtual Environments 17, 29–42 (2008)
6. Sarakoglou, I., et al.: Free to Touch: A Portable Tactile Display For 3D Surface Texture Exploration. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3587–3592 (2006)
7. Van Boven, R.W., Johnson, K.O.: The limit of tactile spatial resolution in humans: grating orientation discrimination at the lip, tongue and finger. Neurology 44, 2361–2366 (1994)

ClearSpace: Mixed Reality Virtual Teamrooms

Alex Hill, Matthew Bonner, and Blair MacIntyre

Georgia Institute of Technology, 85 Fifth Street NW
30308 Atlanta, USA
{ahill,matt.bonner,blair}@gatech.edu

Abstract. We describe ClearSpace, a tool for collaboration between distributed teamrooms that combines components of virtual worlds and mixed presence groupware. This prototype is a starting point for exploring solutions to display and presence disparity by leveraging model-based user representations. We describe our deployed system and a mirroring approach that solves several problems with scaling up ClearBoard style portals to a common virtual space. We also describe techniques for enforcing consistency between heterogeneous virtual and physical contexts through system-managed awareness.

Keywords: Mixed Reality, Distributed Groupware, Mixed Presence Groupware.

1 Introduction

In our global, information-driven economy, distributed workforces have become a fact of life. Highly distributed groups face significant challenges to collaboration and to forming a shared group culture. While issues such as working asynchronously or the loss of physical cues like gestures are key, perhaps the thorniest problem is a distributed workforce's heterogeneity. An in-house user study by Sun Microsystems found that 70% of distributed meetings involved a combination of co-located teamroom users and remote users [1]. Workers may be in offices or on trains, conversing at a table or on a mobile, or in one of a thousand other work contexts. Crafting a shared work context from a potentially infinite number of individual contexts is no easily specified task. Both virtual worlds and Mixed Presence Groupware (MPG) researchers have confronted this problem. Virtual worlds address user heterogeneity by creating a new, virtual world to be shared. MPG systems in contrast carefully track physical cues to bring remote users into an existing workspace, creating a shared "mixed" reality. Our ClearSpace project explores the capacity of a system to blend these the extremes of physically and virtually focused systems, combining virtual worlds and MPG approaches and gracefully handling discrepancies in device configuration to support many heterogeneous users. We have designed a system based on the open-source Project Wonderland virtual world toolkit, which uses a magic mirror technique to mix the physical and virtual work contexts into a single persistent work environment. In our prototype system, we have developed a whiteboard-style application inspired by Ishii's Clearboard [2], with virtual analogues. Rather than using video, we use several tracking systems to map physical motion to articulated avatars in the virtual world.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 333–342, 2011.

Our system has served as a base for exploring solutions to presence and display disparity as described by Tang et al [3]. We are particularly concerned with consistency between physical and virtual contexts, especially user perspectives and the physical or virtual placement of work artifacts within the combined teamroom. We believe realizing consistency on many levels including appearance, affordance and physical to virtual mapping is critical to overcoming presence and display disparities. The next section covers related work and the following section describes our prototype system. Section 4 discusses how a requirement for multiple access points to teamroom documents lead to a mirror-based approach rather than a window approach. Section 5 discusses the management and awareness of workspace documents. Section 6 describes how system managed awareness facilitates scaling across heterogeneous physical to virtual configurations.

2 Related Work

The virtual worlds approach to remote collaboration, exemplified by cAR/PE! and GAZE, minimizes individual physical context by creating a new, shared virtual context [4, 5]. In both systems, video representations of users are placed in a virtual environment or teamroom. GAZE and cAR/PE! also incorporate some degree of head tracking to communicate gaze or control position in the virtual world. While simple tracking can help communicate physical cues, no support for co-located users is provided, and all users may as well be remote. Such enforced homogeneity may overcome certain challenges of creating a shared context, but it does so at the expense of natural co-located interaction. Mixed Presence Groupware (MPG) emphasizes the physical. These systems, such as VIRTUE, Carpeno and VideoArms center on an augmented room or artifact like a table or vertical display that becomes a shared context [6-8]. Sophisticated human tracking systems are used to communicate as many physical cues as truthfully as possible. Rather than bringing users into a new virtual environment, these systems seek to bring remote users into the same physical realm as co-located users. Such specialized environments can be costly, making these systems difficult to scale. Users in a different context, perhaps on a mobile device at a field site, are unable to participate.

Display and presence disparity, described by Tang *et al*, are particularly important to our work [3]. Display disparity refers to the difficulty of sharing contexts or viewpoints. Users may have different sized displays or displays may be at different orientations, making referring to work artifacts cumbersome. Presence disparity refers to the feeling of co-presence that exists between collaborators. Users collaborating through a remote system are limited to whatever tracking or communication signals are supported, which can lead to a distinct lack of presence. A user whose communicative repertoire has been impoverished by their means of accessing the system may feel noticeably 'apart' from other collaborators. Explicit concern with display and presence disparity is a key difference between ClearSpace and some closely related work. Both VIRTUE and Carpeno for example combine interactive table environments with virtual worlds [6, 7]. In either system, one can work with a remote collaborator on a horizontal interactive table, or look 'across the table' at a vertical screen for visual representation of their partners. Our system also combines elements of MPG and virtual worlds, but our focus on display and presence disparity has led to interesting differences in our design. Perhaps the most obvious of these is our use of "mirroring", through which remote users appear to be physically located alongside physical users. Through this approach, we hope to create a vertical display that doesn't form a cognitive 'wall' between collaborators, even without the aid of a conjoined collaborative table.

3 System Configuration

Ishii’s ClearBoard project had the benefit of placing video cameras near the remote user location behind the interaction surface, which also allowed the system to capture interactions with the whiteboard surface. A more conventional placement of a camera at surface depth does not allow the capture of hand interactions with that surface. A further drawback of using video in general is that it highlights differences in the user context such as lighting, clothing, background and even camera resolution and connection bandwidth. In contrast, a model-based representation tends to mask these differences, which aligns well with our approach, which is to reduce display and presence disparity through consistent representations for all users. We understand that any model-based representation we develop will fail to capture the subtlety and range of video. However, we believe that the increasing use of human tracking technology (i.e. Microsoft Kinect) suggests that these approaches will eventually capture as much if not more fidelity than video alone. Furthermore, we feel that starting from a model-based approach allows us to consider a much greater range of scenarios than normally apply when using video alone.

Before engaging in an analysis of how to best merge virtual worlds with interactive whiteboards we had to determine a hardware and software configuration that met some minimum requirements for capturing the user and their input. Feeling that tracking the head and hands brings the highest marginal value in representing overall physical embodiment, we chose a combination of vision-based head tracking and hand tracking in front of a large LCD panel. The prototype we developed consists of two 65” Sharp displays each mounted with multitouch overlays from NextWindow and Logitech QuickCam Pro 9000 cameras. For head tracking we used the FaceAPI library from SeeingMachines, which tracks the head location and orientation of a single user in three dimensions along with lips and eyebrows. Another commercial system, the ZCam from 3DV, had been demonstrated to track hands and recognize some simple gestures but became unavailable after Microsoft Corporation purchased it. As a fallback, we turned to an OptiTrack fiducial tracker with multiple cameras (6) focused on the interaction space in front of the LCD display. We deployed our virtual teamspaces using Sun Microsystems Wonderland open source Java library, which features full persistence and module-based customization. The base Wonderland system includes several features that support collaboration such as an interactive SVG whiteboard module and stereo directionalized voice conferencing. We developed a VRPN client module for acquiring head and hand tracking data, and customized the avatar character class to perform arm inverse kinematics from hand position. We modified the SVG whiteboard module to include free-hand drawing and interactive client updates during gestures. We also added features that let us store and load documents using a WebDAV repository. We also created a Metadata module for tagging documents and a ProjectManager module that lets the user choose from a set of associated documents from within a HUD view.

4 Creating a Shared Collaborative Space Using Portals

One of our main goals was to allow scalability and flexibility in how the system is configured and used. Real world use of whiteboards in teamrooms provides the flexibility to leave content on multiple boards over time and to use any of those surfaces at any time. Like many other collaboration systems, the ClearBoard design creates a portal to the remote site on the other side of a display surface. While most systems use the display as a window into the virtual world we realized that this limits ability to scale to multiple portals. In order to implement an interactive whiteboard across the physical/virtual divide, the rendering into the virtual space has to be flipped across the vertical display plane. Using a window paradigm, a second portal on an adjacent wall in the room adds a second viewpoint into the virtual world from beyond another similarly placed portal in that space (Fig. 1a). The local portal user and a virtual collaborator can both turn in the same direction and transition their collaboration onto the second portal in a natural manner.

Unfortunately, using a window portal approach creates some problems. Given physical user P looking into a portal with another to their right, the arrangement of the virtual world will appear to be reversed with the second portal residing to the left of users in the virtual world. When the scene is later viewed through a virtual world browser (VWB), this inconsistency may interfere with workflow or even create confusion. A second issue arises when localized sound sources are added to the configuration. Sound is a critical part of distributed collaborations and directional sound has been shown to facilitate discriminating individual speakers amongst other competing voices [9]. If the sound of user V is directional, then its source will have to make an abrupt transition from one location to another as user P turns to the other view (Fig. 1a). Finally, having multiple viewpoints into the virtual space begs the following question: If a third user V_2 joins the collaboration through a VWB, do they find user P on the same side of the whiteboard surface as user V_1? If the user P viewpoint is indeed looking into the teamspace from outside then on which side should V_2 join the collaboration? If user P is indeed inside the teamspace with V_1 and V_2 his will inevitably lead to confusion as V_2 may believe they are standing next to user P when in fact they are across from him.

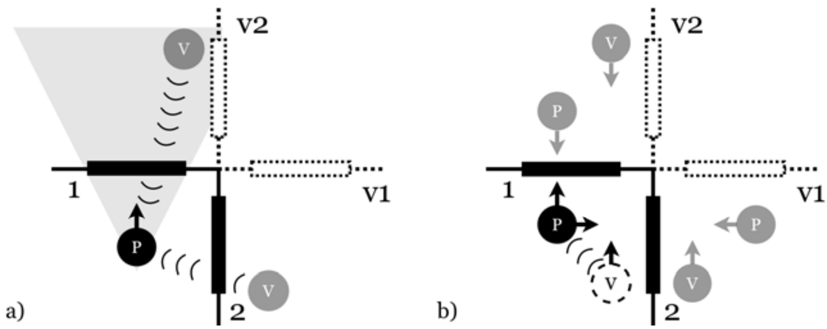


Fig. 1. In a), a physical user P looks through window portal 1 at virtual user V for whom portal 2 appears to be on their left. If P turns to look through window portal 2 the audio for V must transition. In b), mirror portals avoid these issues since user P acknowledges that the portal is the right of V. As P turns the audio source for V inside the room remains consistent.

The solution to these problems lies in acknowledging that transitions between physical portal and similarly aligned virtual portal views behave in much the same way as collaborating in front of mirrors. The main implication being that a mirror view reflects back the virtual image of user P (Fig. 1b). Making this subtle change suggests that user P is actually sharing space with user V. Instead of acknowledging separate local space and remote spaces, the two spaces are fused into a common shared space. There is already some empirical support that virtual users respect personal space in much the same way as physical users do [10]. One can easily imagine a scenario where P and V_1 are collaborating and V_2 approaches from the rear. When P and V_1 see themselves and the approaching user V_2 sharing the same space they will naturally move aside, acknowledging the presence of user V_2 in their own space. A mirrored paradigm resolves questions about where each collaborator resides because all are collaborating in the same virtual teamspace side-by-side. A mirror paradigm allows gaze and gestures between users to work in a natural manner without software intervention. This approach also resolves discontinuities with directional sound. When user P acknowledges that V is standing to his side it becomes possible to locate the sound source of the collaborator within the same room (Fig. 1b). When turning to look into another portal, the sound location of V within the room stays consistent with their visual location in the other portal.

The foreshortening that happens close to a virtual camera makes showing activity within a virtual teamspace impractical with window portals. However, they can increase awareness of the virtual context for user P by creating views into the space surrounding the teamspace through virtual doors and windows. Analogous to their function in physical spaces, these apertures give feedback to virtual users indicating when they are visible to physical teamroom users. While head tracking can provide perspective corrected views in a mirror portal, window views are best rendered using a fixed viewpoint since they will be viewed from disparate viewpoints within the room. This general rule enforces another of our principles for the ClearSpace design, that user distance from portals should reduce the granularity of the information they provide.

5 Managing Teamroom Work Artifacts

Documents produced by portals include whiteboard drawings, affinity diagrams and introduced PDF documents, images and 3D models. A VWB user can ostensibly peel a finished document off of a portal and place it arbitrarily within the teamspace. Moreover, these VWB users can also travel arbitrarily to documents in order to access them. This raises an important question: Should all documents be portals in themselves? If each document is a portal in itself then two VWB users can conceivably collaborate on a document away from the physical portal. If physical portal users are prevented from joining in on the collaboration, then this lack of consistency in affordance will contribute to presence disparity. To avoid this, VWB users can travel to, access and manipulate documents in any location privately, but all collaborations must happen through portals. When users want to make their actions pubic and available for collaboration, they have to move a document onto a portal. This requires the user travel to a nearby portal, determine if another user is currently using it and coordinate placing a copy of the document into the portal. A VWB user accessing a document can also bring the document into a heads up display (HUD) mode. In HUD mode, the user has the option to see users publicly manipulating the document appear behind it. The decision to move to a portal can start by switching viewpoint to that of the portal that created the document. From there, the presence of other users, their activity and negotiation for portal space can begin.

A well-established design principle of Computer Supported Collaborative Work (CSCW) is the need to maintain awareness of collaborator activities. In many collaborative workspaces global actions are inherently visible to other users, while in virtual environments a common shared viewpoint cannot be guaranteed. Remaining aware of the viewpoint of virtual users via their location and head orientation helps in this regard. Some virtual environments such as Second Life do not enforce a linkage between user embodiment and viewpoint. In an effort to ensure consistency of both affordance and appearance, user viewpoint should always be tied to their embodiment. This is no longer the case when a user brings a document into a HUD view and some visual feedback to users becomes necessary to maintain awareness. In this situation, the VWB user can be automatically moved into a position in front of the associated document and returned to their position when the interaction ends. A semi-transparent trace avatar can also remain at the original location to avoid discontinuities in user awareness. This same technique can also be used when a VWB user decides to take HUD interactions to a portal for public collaboration. If the user decides to stay in front of the document or portal their trace avatar can move to join them there, making other users aware of the action.

Although using a virtual world as a document repository makes user actions with those documents visible, forcing all users to travel to find and access documents is impractical. A useful HUD addition is an overview mode that includes thumbnails of associated work artifacts created by the same portal. When a VWB user employs this overview HUD they are no longer just interacting with a single document and a trace visualization of the overview HUD in front of the user is used. Access rights to user information determine the degree to which users can actually see the details of this behavior. To ensure consistency, this same overview mode is available to physical portal users as well. When a portal user takes his interactions private their view of the portal may change but those of other users at the portal remain the same. Portal users become aware of private actions of other users through a change in their appearance to semi-transparent and a trace of their activity. This trace can appear localized to an area around the virtual user in order to reduce any conflict between it and the document visible in front of the physical portal user.

A key ingredient in teamroom activities is the placement of documents in locations throughout the space. When a new document is introduced into the teamspace, a VWB user can ostensibly place that document anywhere in the environment. When introduced via a portal, either by whiteboarding or some other means, it is appropriate to place the document around or beyond that portal. Documents in the virtual world are associated with the portals that introduce them and should benefit from a visual indication by either restricting their location to a portal wall or visually indicated area around the portal. Physical portal users are somewhat more restricted in their ability to arrange documents within the scene. A simple heuristic for blank document creation is to position finished documents on the wall in sequence from left to right. When more explicit control over document position is required, a fixed third person view of the portal is appropriate. Using the same fixed view of the portal between physical and VWB users provides consistency. The fixed view around the portal should let users scale less prominent documents to a smaller size. This action can either result in scaling the document or only the appearance of scaling by pushing documents back in depth. Algorithmically adding depth to documents should be avoided if it begins to compromise the ability of teamroom users to appreciate document arrangement when viewed later through a VWB.

A portal view using a mirror paradigm will only include documents arranged behind the portal within its viewing frustum. And, a third person viewpoint will not facilitate arranging documents behind the user. One solution to this problem is to project a window portal view from a fixed perspective onto the wall containing the actual portal. This display around the portal creates a focus-and-context style display and gives the user arranging documents immediate feedback about their actions through peripheral vision. With hand tracking and some simple gesture recognition in front of the portal, a user can move documents off of the portal space and into position on the wall. Another way to make documents around a portal visible is to show the third person viewpoint instead of the mirrored view of the wall behind the user. Regardless of what is shown beyond the screen, some form of mode switching will be necessary to disambiguate between actions directed at the portal from those directed at content beyond it. A third person view of content on the wall around the user can also be created in a physical room by placing a mirrored wall at the back of the room. It may provide more spatial consistency to the user to use a virtual mirrored rear wall to construct the third person view that includes yet another view of the user within it.

6 Exploiting Model-Based Representations to Manage Awareness

One of the most significant advantages of a model-based user representation is the flexibility it affords. Procedural manipulation of user representation can be used to exaggerate awareness of user behaviors by amplifying their rendered signal [11]. This can also be extended to synthesizing behaviors such as gaze that have not been captured but are inferred from user behavior [12]. An example of how this might be exploited in a mirror portal setup involves awareness of gaze between two remote users. Given P side-by-side with virtual collaborators V_1 and V_2, when V_1 addresses V_2 he will only turn his head slightly to make eye contact with that user (Fig. 2a). When interpreted by the system, the awareness of this action to P can be enhanced visually by showing V_1 actually turning to face V_2. Once the underlying system is managing the gaze and gestures of users it also becomes feasible to have the system manipulate their apparent location. The space in front of portals will undoubtedly get crowded with the addition of multiple collaborators and the inclusion of mirrored avatars. Instead of having users explicitly coordinate their use of portal space, implicit heuristics can manage proxies of each user differently in each portal. Like some voice conferencing support systems, those users actively engaging in conversation or whiteboard activities can be represented near the front while those not participating are moved to a location around the periphery (Fig. 2b).

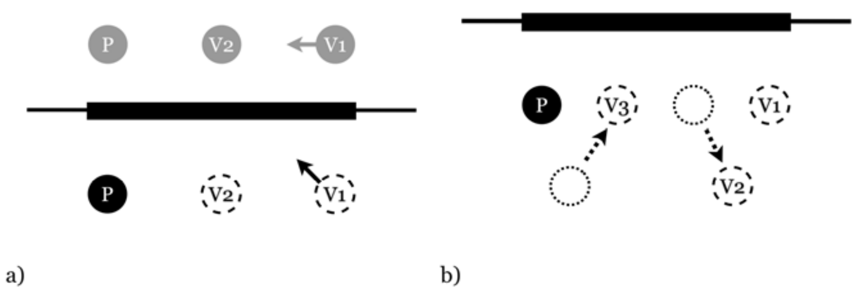


Fig. 2. In a), user V_1 turns slightly to look at V_2 but his representation turns to completely face V_2. In b), the representation of a less active V_2 moves to the periphery while a more active V_3 moves forward

The straightforward way to ensure a consistent relationship between physical portals and virtual portals is for each to share the same size and relative location. More realistically, there will be mismatches between not only different physical teamroom setups but also between physical portals and their virtual equivalent. When the locations of portals do not align, the transformation of user position into those coordinate systems will create discontinuities as the user moves between portals. A model-based representation allows the system to leverage blending to generate consistent behaviors as users transition between different frames of reference. A simple heuristic that can be used in this context involves using the relative head orientation towards each portal. Following this approach, a user facing a portal would see virtual user V in the appropriate position relative to that portal (Fig. 3a). However, as user P turns to address another portal, the avatar of user V would appear to move as necessary to assume their appropriate position relative to the portal in view (Fig. 3b).

A more common discrepancy will be relative size, which results not only from physical portals of varying sizes but also from mobile devices. Transforming user behaviors between a small portal and a larger portal simply means applying them to an avatar sized appropriately for the larger context (Fig. 3c). Both user A interacting with a laptop and user B at a smaller portal will have their movement scaled up to fit their presentation. For laptop user A, gaze and mouse activity become a combination of gaze and body movement in the remote portal. Similar heuristics can also be applied to allow portals to pan and zoom into documents. The main result is that portal users no longer share the same viewing frustum of the portal or the users beyond it. When a remote user zooms into a document, the reduced viewing frustum is highlighted by a rectangle on the remote portal (Fig. 3c). When combined with their mirror reflection, this rectangle gives users visual feedback about their visibility to the remote user. Larger portals do not have to result in larger documents as sub-parts of whiteboard space can be saved off as individual documents. As portals increase in size to that of a wall, their effective work area becomes bounded by the size of the portal in the virtual teamspace and the viewing frustums of remote collaborators.

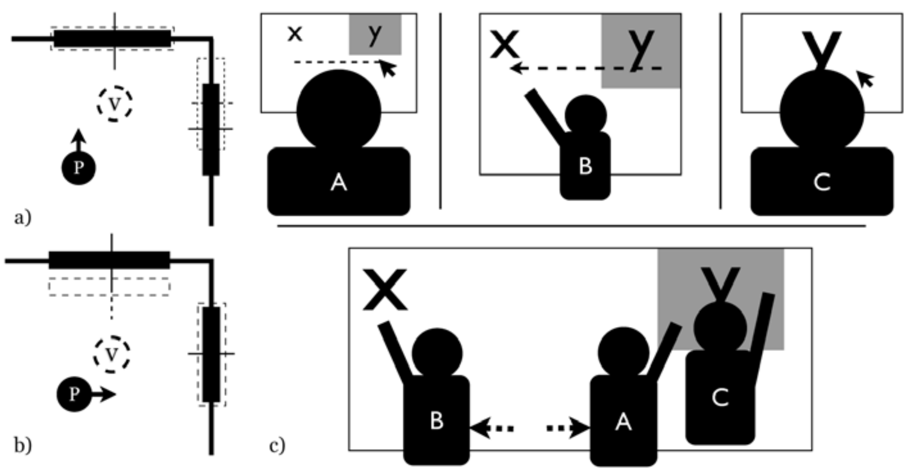


Fig. 3. In a), user V appears in the actual position relative to the front portal and in b), appears relative to the side portal when P faces it. In c), laptop user A and remote user B both appear at the same scale in a larger portal. A reduced frustum of user C appears as a rectangle.

7 Discussion and Future Work

In this project, we developed a ClearBoard inspired distributed collaboration system based on the Wonderland Java library. Instead of using video, we used model-based representation in an effort to fully explore the potential of using mixed reality techniques and system-managed awareness to decrease display and presence disparity. In contrast to using a window into the virtual environment we found that a mirrored paradigm solved a number of problems with scalability and flexibility. Our analysis suggests that managing user awareness will help handle discrepancies in portal size and location. In addition to executing many of the unimplemented system features described here, our future work will include replacing our OptiTrack fiducial tracking system with one or more Microsoft Kinect controllers.

Acknowledgments. We would like to thank Sun Microsystems for their financial support, Jacob Schiefer, Aneesh Tarun and the Steelcase WorkSpace Futures team for their generous time spent inside Wonderland evaluating ClearSpace concepts.

References

1. Yankelovich, N., Simpson, N., Kaplan, J., Provino, J.: Porta-person: telepresence for the connected conference room. In: CHI 2007 Extended Abstracts on Human Factors in Computing Systems, pp. 2789–2794. ACM, New York (2007)
2. Ishii, H., Kobayashi, M.: ClearBoard. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, Monterey, California, United States, pp. 525–532 (1992)
3. Tang, A., Boyle, M., Greenberg, S.: Understanding and mitigating display and presence disparity in mixed presence groupware. J. Res. and Practice in Inf. Tech. 37
4. Regenbrecht, H., Ott, C., Wagner, M., Lum, T., Kohler, P., Wilke, W., Mueller, E.: An Augmented Virtuality Approach to 3D Videoconferencing. In: Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality, p. 290. IEEE Computer Society, Washington, DC, USA (2003)
5. Vertegaal, R.: The GAZE groupware system: mediating joint attention in multiparty communication and collaboration. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 294–301. ACM, New York (1999)
6. Kauff, P., Schreer, O.: An immersive 3D video-conferencing system using shared virtual team user environments. In: Proceedings of the 4th International Conference on Collaborative Virtual Environments, pp. 105–112. ACM, New York (2002)
7. Regenbrecht, H., Haller, M., Hauber, J., Billinghurst, M.: Carpeno: interfacing remote collaborative virtual environments with table-top interaction. Virtual Reality 10, 95–107 (2006)
8. Tang, A., Neustaedter, C., Greenberg, S.: VideoArms: Embodiments for Mixed Presence Groupware. In: Bryan-Kinns, N., Blanford, A., Curzon, P., Nigay, L. (eds.) People and Computers XX — Engage, pp. 85–102. Springer, London (2007)
9. Yankelovich, N., Walker, W., Roberts, P., Wessler, M., Kaplan, J., Provino, J.: Meeting central: making distributed meetings more effective. In: Proc. 2004 ACM conference on Computer Supported Cooperative Work, pp. 419–428. ACM, New York (2004)
10. Miwa, Y., Ishibiki, C.: Shadow communication: system for embodied interaction with remote partners. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, pp. 467–476. ACM, New York (2004)
11. Deng, Z., Ma, X.: Perceptually guided expressive facial animation. In: Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, Eurographics Association, Aire-la-Ville, Switzerland, pp. 67–76 (2008)
12. Steptoe, W., Oyekoya, O., Murgia, A., Wolff, R., Rae, J., Guimaraes, E., Roberts, D., Steed, A.: Eye Tracking for Avatar Eye Gaze Control During Object-Focused Multiparty Interaction in Immersive Collaborative Virtual Environments. In: Virtual Reality Conference, pp. 83–90. IEEE Computer Society, Los Alamitos (2009)

Mesh Deformations in X3D via CUDA with Freeform Deformation Lattices

Yvonne Jung, Holger Graf, Johannes Behr, and Arjan Kuijper

Fraunhofer IGD, Fraunhoferstraße 5, 64283 Darmstadt, Germany
{yvonne.jung,holger.graf,johannes.behr,
arjan.kuijper}@igd.fraunhofer.de

Abstract. In this paper we present a GPU-accelerated implementation of the well-known freeform deformation algorithm to allow for deformable objects within fully interactive virtual environments. We furthermore outline how our real-time deformation approach can be integrated into the X3D standard for more accessibility of the proposed methods. The presented technique can be used to deform complex detailed geometries without pre-processing the mesh by simply generating a lattice around the model. The local deformation is then computed for this lattice instead of the complex geometry, which efficiently can be carried out on the GPU using CUDA.

Keywords: Deformable objects, real-time simulation, FFD, CUDA, X3D.

1 Introduction

In this paper we present a hardware accelerated implementation of the freeform deformation (FFD) algorithm to allow for fully deformable objects in interactive virtual environments. The presented method can be used to deform complex detailed geometries without pre-processing the mesh by simply generating a lattice around the model. The local deformation is then computed for this lattice instead of the complex geometry, which efficiently is carried out on the GPU using CUDA. Freeform deformation is a common technique in computer graphics and image processing. In the field of medical imaging for instance, the grid-based FFD is used for registration approaches [2, 3], today's DCC tools (like e.g. 3ds Max [14]) utilize it as another modeling method, whereas we discuss the FFD in the field of object animation [4].

Despite the advancements in simulating deformable objects such as cloth [12, 15], hair [8, 17], or biological tissues [10, 11, 19] – not least because of the proliferation of multi-core CPUs and the rapid GPU development cycles [16] – it is still rather difficult, if possible at all, to integrate such techniques into interactive 3D applications like in Virtual or Mixed Reality environments [17, 20]. Hence, for ease of access, in this paper we will first discuss, how we have integrated this deformation method into X3D [9], an open ISO standard that not only defines a 3D interchange format, but which can also be used as a declarative application description language. The integration is achieved by means of a new physics material node type that is designed as a part of an object's general appearance. This idea is similar to the approach recently presented in [7], where the authors proposed an additional X3D node set that

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 343–351, 2011.

e.g. contains a special physics node for defining the physical properties associated with a shape node's geometry. Although these extensions originally were only intended for haptic rendering, they have the potential to support more dynamic and realistic object behaviors.

Next, we present and compare our implementation with focus on the various CUDA kernels used, while utilizing the OpenGL – CUDA interoperability as described in [5] for maximum performance. Here, we have basically implemented two different methods: the first one is based on the Bernstein polynomials as proposed in the original algorithm [1], while the other one utilizes a mass-spring-system that acts on the control points as described in [8]. Although the first method is not physically-based, it is very fast, easy to implement even with CUDA, and allows achieving useful and interesting deformation effects during runtime, like elastic objects of arbitrary shapes that bounce in a plausible manner. Due to the inherent connectivity of the mass (or control) points in a mass-spring-system, the latter method is more intricate to implement on the GPU due to its SIMD architecture, but allows the application developer to design highly dynamic scenes.

To further optimize performance, we propose to use two different CUDA kernels. For objects with many vertices we can reach a full occupancy of all multiprocessors on the GPU. Therefore, vertex deformations are computed per thread: each thread needs all control points and the corresponding vertex. However, for low-poly objects only a few multiprocessors are used. Therefore, we will present our second approach based on parallel reduction. After that, we will present our runtime results and discuss some application examples, before we finally conclude the paper.

2 Related Work

In computer science there exist different methods to simulate deformable objects. Usually one distinguishes between physically-based deformation methods, like the finite element method (FEM), and geometrical ones. The latter only change the shape of polygonal models and are suitable as a tool for intuitive deformation tasks such as in CAD- or modeling-packages [14]. A well-known geometrical method is the grid-based free-form deformation. The original FFD method already was published by Sederberg and Perry 25 years ago [1], and is basically a geometric spline deformation based on the trivariate Bernstein polynomials [1, 18], which can be manipulated through a 3D lattice of control points, in which the geometries to be deformed are embedded. The idea is exemplarily visualized in Fig. 1 with a linear deformation.

However, other deformation schemes are possible, too, such as the mass-spring-based simulation presented in [8], which combines this indirect deformation method with a more physically-based type of mesh deformation to animate complex hairstyles in real-time. Similarly, the SOFA framework for medical simulations is presented in [10]. Along with other grid-based deformation schemes a mass-spring-system is utilized for object deformations, which in a prototypical implementation is further accelerated with CUDA [6] by exploiting the highly parallel nature of modern graphics hardware. In [11], researchers from the same group lately also presented a method for simulating detailed geometric models using coarse grids by taking into account the topology and physical material of the embedded geometry to overcome problems like disconnected parts (with possibly different material properties) that are located within the same grid cell.

Generally, to achieve an animated deformation using physically-based methods, a simulation system needs to be advanced in time. Therefore, in [11] the implicit Euler integration scheme [12] is used to solve the motion equation for the given time step. The type of differential equation to be solved depends on the type of motion equation. Newtonian motion is based on Newton's second law of motion (i.e., $F = m \cdot a$) and can be solved by integrating an ordinary differential equation. The implementation is relatively easy because of the discretization to a finite number of mass points, and the integration is independent for each such particle.

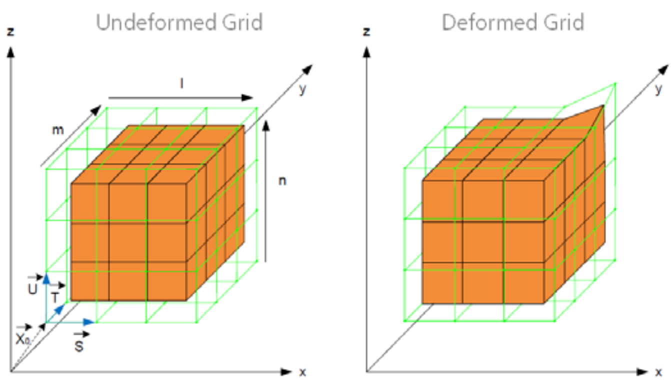


Fig. 1. Basic principle of grid deformation: on the left, exemplarily a cube embedded in a surrounding 3D lattice is shown, and on the right, the deformed grid/ cube is visualized

As already mentioned in the introduction, it is still rather difficult to integrate more advanced deformation techniques into interactive 3D applications like in VR or MR environments. Therefore, in [7] the authors proposed an additional X3D node set that, amongst others, contains a special physics node type for defining the tangible physical properties that are associated with a shape node's geometry (analogously to the X3D appearance node that is responsible for the optical properties of a certain geometry). Likewise, this physics node can be parameterized with various child nodes for specifying friction and so on. However, these extensions were only intended for haptic rendering, though they have the potential to support more realistic and physically-based object behaviors, e.g. in case of collisions with other scene objects. Similar in spirit, the VR/AR framework Instant Reality [13], which utilizes X3D as application description language, provides a "simulator" component that allows modeling simulation systems (e.g. for flow fields, particle dynamics or hair).

In this regard, X3D already provides a "rigid body physics" component, which describes how to model rigid bodies and their interactions, and a "particle systems" component, which specifies how to model particles and their interactions through the application of basic physics principles to affect the resulting motion [9]. However, simulated deformable objects are still not part of the current X3D standard.

3 Concept and Implementation

In this section, we first describe the high-level concept, i.e. how we have integrated dynamic mesh deformations into X3D [9] with a new set of nodes for interactively controlling and parameterizing the desired deformations. After that, we explain our CUDA-based implementation to achieve better performance.

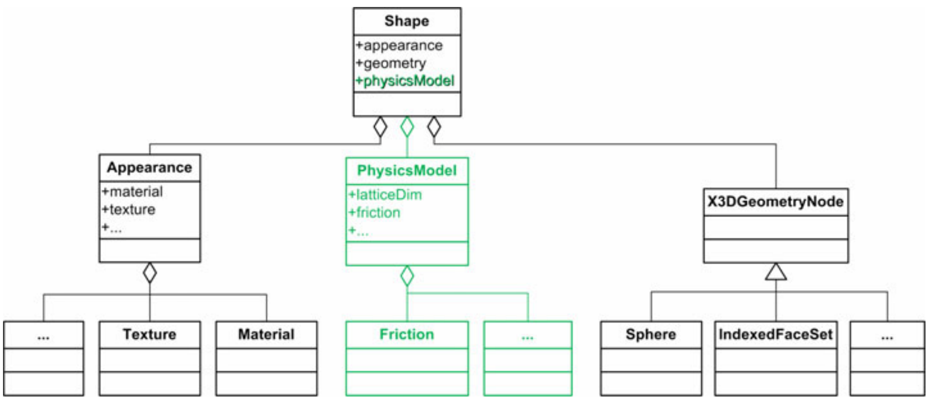


Fig. 2. Class diagram of the X3D *Shape* node: child nodes that are part of the X3D standard are shown in black, whereas the proposed new *PhysicsModel* node set is depicted in green

3.1 Proposed X3D Node Extensions

As already mentioned, our extension proposal basically follows the concepts as described by Wei et al. [7], though their focus was on enriching the standard X3D *Shape* node (which already aggregates the mesh geometry and appearance or optical properties – see Fig. 2) by function-defined geometry and tangible physical properties to allow for haptic exploration. Similar in spirit, with the *ParticleSystem* the X3D "particle systems" component defines a special type of shape node that specifies a complete particle system, since appearance and geometry both need to be controlled by the simulation to create plausible particle effects.

In contrast to the special case of particle physics [9], and following [7], we extend the standard shape node with a field 'physicsModel' for defining the object's physical properties in a composable way (as shown in the middle of Fig. 2), in order to support effects such as friction, restitution, or gravity and other external forces etc., which are not only useful for specifying deformation parameters but also for rigid body physics. If none of the *PhysicsModel*'s child nodes are specified but a lattice dimension is set and some force is applied, the system defaults to geometric deformation.

3.2 GPU-Based Implementation

We use OpenGL and the interoperability with CUDA to implement the FFD on the GPU. The first step is to initialize OpenGL and load the mesh to a so-called VBO (Vertex Buffer Object), which conceptually is an array of bytes that can handle the upload of vertex attributes such as vertex positions, normals, colors and texture coordinates to the device memory of the graphics card. The deformation is then done within the rendering loop. In each iteration, the deformation is computed from the original object and the result is stored in a separate VBO, which we call *Deformed Mesh VBO*. Afterwards, the deformed mesh is used for display.

The lattice dimension is determined by the user manually via the SFVec3f field 'latticeDim' of the proposed *PhysicsModel* node (see greenish middle part of Fig. 2). All relevant lattice data is stored in a separate VBO. Based on the number of control points in each direction, the resulting object lattice is positioned automatically. The control points are arranged in equidistant steps (compare Fig. 1).

Moving a control point leads in turn to a deformation of the attached object (note that the mesh indices do not change during the deformation calculation). Therefore, the deformation function inside the rendering loop is called and the new positions of the mesh vertices are calculated, whereas this calculation depends on the chosen type of deformation. Currently, two deformation methods are provided: a geometrical one based on Bernstein polynomials [1] (outlined in sections 3.3 and 3.4) and a more physically-based that follows a mass-spring-system approach (see section 3.5).

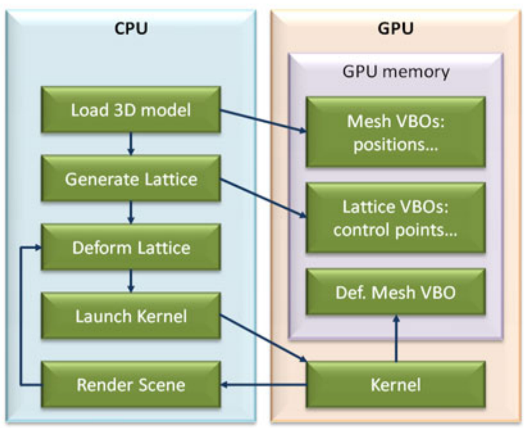


Fig. 3. Illustration of implementation concept (with CPU and GPU part)

In Fig. 3, the overall concept of the FFD implementation is coarsely sketched. Here, deforming the lattice means, that e.g. the user interactively changes the position of the control points (or respectively deforms the grid with a force F). Afterwards the CUDA kernel computes the FFD including the distance d from each point p to the surrounding control points. The weighted translation vector f is added to p and the result is then stored in the Deformed Mesh VBO.

The FFD itself is realized using CUDA. Before the data in the VBOs is usable, it has to be registered once as a CUDA resource. To gain access to the VBO data, a pointer to the memory location is needed. This is done by mapping the VBO resource to the CUDA address space. The result of the mapping is a pointer to the relevant data which can now be used in CUDA kernels. A more detailed description of the OpenGL – CUDA interoperability can be found in [5] on page 38 – 40.

To further optimize performance we have implemented two different CUDA kernel approaches. One kernel is optimized for a low amount of vertices per object (about 500) and the other kernel is adjusted for objects with thousands of vertices.

3.3 CUDA Kernel for Complex Objects

For more complex high-resolution models we can easily reach a full occupancy of all multiprocessors (MP). Therefore, we compute the deformation of one vertex in one thread. Each thread needs all control points and the corresponding vertices. As mentioned in the previous section, we first focus on geometric mesh deformation methods, since these are much easier to parallelize. Hence, for each control point the Bernstein polynomial will be computed and therefore it is important, that the vertices as well as the control points are copied to shared memory. We choose a block size of 16 × 16. The first step is that each thread uploads one vertex to the shared memory. The vertices will be copied in a coalescing way, because 16 threads in a row copy vertices that are stored consecutively in the global memory.

In our approach the control points will not be copied to shared memory, because this will limit the lattice dimension and the occupancy of an MP can decrease. With a device of compute capability 1.3 (e.g. an NVidia GeForce GTX 295 graphics card) 1024 threads $\boldsymbol{t}$ can be hold by one multiprocessor. In our case four blocks will be assigned to one MP. The following equation shows, that 12 kB of the shared memory $\boldsymbol{s}$ are used by the vertex data (where one vertex consists of three float values):

$$\boldsymbol{s} = \boldsymbol{t} \cdot 3 \cdot sizeof(float) = 1024 \cdot 3 \cdot 4\ Byte = 12\ kB$$

The shared memory thus has a total amount of 16 kB. If a full occupancy of one multiprocessor is desired, then the maximum size of the lattice is 85 control points. To reduce the latency of the copy of the control points from the global memory, every thread loads identical control point data. This leads to broadcast memory accesses where the copy is fully parallelized. After copying the data, the FFD will be computed and the result is stored at the address of the Deformed Mesh VBO.

3.4 CUDA Kernel for Simple Objects

When an object with only a few vertices shall be deformed, it is not ideal to use the CUDA kernel for complex objects. Assume that we have an object with 512 vertices. Then there will be only two blocks with 16 × 16 threads and only two multiprocessors of the GPU will be used. To alleviate this issue we compute the deformation of one vertex in one block. Each thread in a block will compute one “sub” deformation. This means that a thread computes one Bernstein polynomial of one control point. Therefore, each thread uploads a control point to shared memory and computes a sub deformation. This sub deformation must then be summed up to get the complete deformation. This is done with the help of parallel reduction. To do the parallel reduction, similar to the FFT algorithm in signal processing, the thread count of one block must be the next power of two. A disadvantage of this technique is that the lattice can only have as much control points as the maximum count of threads per block, but in general the lattice for small objects has much less control points.

3.5 Dynamics Simulation

In order to simulate elastic objects, the other method is based on a mass-spring-system that acts on the lattice, as originally proposed in [8] for hair simulation. In our first prototypical implementation we simply identify Newton's second law of motion with Hooke's law of elasticity (i.e., $F = D \cdot \Delta l$). Then, the forces between all control points are updated in each frame and applied to the control points. To parallelize this approach all possible forces between the control points are independently computed by following the ping-pong "rendering" approach (via a helper buffer) as known from GPGPU programming. Afterwards, each control point is handled by a thread that computes the movement, which results in the final deformation.

4 Discussion and Results

We have implemented our FFD approach using CUDA and C++ and prototypically integrated the proposed node set into the MR framework Instant Reality [14]. First, we compared the computation time on the CPU with both approaches for simple and complex objects with a different count of vertices for a wireframe goblet model. In Table 1 the results are shown – all measurements are done with a lattice of $3 \times 3 \times 3$. For a small number of vertices there is not much speed up. If the object has less than 5,000 vertices, a slight speed up can be achieved when using the CUDA kernel for simple objects. If the number of vertices reaches 10,000 vertices and more, the computation time of the CPU is very high. With the CUDA kernel for complex objects we can achieve a considerably smaller computation time of the FFD.

Table 1. Runtime comparison of both CUDA kernels vs. computation time on CPU

#vertices	CPU time/ ms	GPU time/ ms (kernel f. simple obj.)	GPU time/ ms (kernel f. complex obj.)
502	7.5	4.0	15.1
3099	47.6	5.2	15.4
34,834	526.0	39.5	15.9

Fig. 4 shows the deformation of two different 3D objects: a simple point cloud given as X3D *PointSet* node (left) and a dino model given as *IndexedFaceSet* nodes (right). The grid's control points are visualized as blue spheres, which for testing purposes can be moved to trigger the deformation process. Table 2 shows the runtime results obtained from both test sets shown in Fig. 4. For comparison, a different hardware setup was chosen here. Instead of the Core2 Quad CPU @ 2.66 GHz (where obviously only one core was active) and a GPU with compute capability 1.3 – as used for the benchmarks in Table 1 – a single core CPU and a caps 1.0 GPU was used.

Table 2. Benchmark (Pentium IV @ 3.2 GHz CPU with GeForce 8800 GTX GPU)

Model	#vertices	latticeDim	CPU time/ ms	GPU time/ ms
Point cloud	5,000	$3 \times 3 \times 3$	15	2.7
´´	10,000	$3 \times 3 \times 3$	47	5.4
´´	10,000	$4 \times 4 \times 4$	94	10.7
´´	10,000	$5 \times 5 \times 5$	203	34.5
´´	10,000	$6 \times 6 \times 6$	359	162.8
´´	50,000	$3 \times 3 \times 3$	219	26.9
Dino	8,145	$7 \times 7 \times 7$	484	207.0

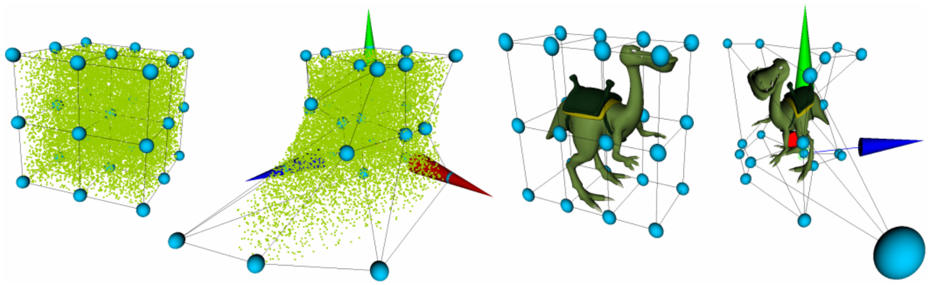


Fig. 4. Left: *PointSet* node with 15,000 test points and 3 × 3 × 3 control points (visualized as blue spheres). Right: Dino model with 2 × 3 × 4 control points (the three colored cones in the deformed image denote manipulation handles). Both are shown undeformed and deformed.

5 Conclusions

In this paper we have presented a hardware accelerated implementation of the freeform deformation algorithm to allow for fully deformable objects within interactive virtual environments. Moreover, we have described how an integration of the presented deformation approach into the current X3D standard can look like to improve the accessibility of the proposed methods also application developers that are no experts in this area. These techniques can be used to easily deform complex geometries without pre-processing the mesh by automatically generating a lattice around the model. The local deformation is then computed for this grid instead of the original geometry, which efficiently can be done on the GPU using CUDA.

Another benefit is the generalization of physical properties that are inherently associated with a certain 3D object, which can also be used to define the respective parameters for both already existing X3D physics component. However, a drawback of our CUDA-based implementation is that high-end graphics hardware (i.e. a modern Shader Model 4.0/ 5.0 GPU) is required.

References

1. Sederberg, T.W., Parry, S.R.: Free-Form Deformation of Solid Geometric Models. In: ACM SIGGRAPH 1986, vol. 20, pp. 151–160 (1986)
2. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. IEEE Transactions on Medical Imaging 18, 712–721 (1999)
3. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S.: Fast Free-Form Deformation using Graphics Processing Units. Computer Methods and Programs in Biomedicine (2009)
4. Faloutsos, P., van de Panne, M., Terzopoulos, D.: Dynamic Free-Form Deformations for Animation Synthesis. IEEE Transactions on Vis. and Comp. Graphics 3 (1997)
5. NVIDIA, CUDA Programming Guide, Edition 3.0 (2010)
6. Sanders, J., Kandrot, E.: CUDA by example: an introduction to general-purpose GPU programming. Addison Wesley, Reading (2010)
7. Wei, L., Sourin, A., Stocker, H.: Function-based haptic collaboration in X3D. In: Proc. of Web3D 2009, pp. 15–23. ACM, New York (2009)
8. Volino, P., Magnenat-Thalmann, N.: Animating complex hairstyles in real-time. In: VRST 2004: ACM Symposium on VR Software & Technology, pp. 41–48. ACM, NY (2004)
9. Web3D Consortium. X3D (2008), `http://www.web3d.org/x3d/specifications/`
10. Allard, J., Cotin, S., Faure, F., Bensoussan, P.-J., Poyer, F., Duriez, C., Delingette, H., Grisoni, L.: SOFA – an Open Source Framework for Medical Simulation. In: Medicine Meets Virtual Reality (MMVR 15), Long Beach, USA, pp. 13–18 (2007)
11. Nesme, M., Kry, P.G., Jerábková, L., Faure, F.: Preserving topology and elasticity for embedded deformable models. In: SIGGRAPH 2009 Papers, pp. 1–9. ACM, New York(2009)
12. Witkin, A., Baraff, D.: Large steps in cloth simulation. Computer Graphics 32, 43–54 (1998)
13. Fraunhofer IGD, Instant Reality framework (2011), `http://www.instantreality.org/`
14. Autodesk, 3ds Max (2011), `http://area.autodesk.com/3dsmax2011/features`
15. Knuth, M., Kohlhammer, J., Kuijper, A.: Embedding Hierachical Deformation within a Realtime Scene Graph – A Simple Approach for Embedding GPU-based Realtime Deformations using Trilinear Transformations Embedded in a Scene Graph. In: GRAPP 2010, pp. 246–253 (2010)
16. Binotto, A.P.D., Daniel, C., Weber, D., Kuijper, A., Stork, A., Pereira, C., Fellner, D.W.: Iterative SLE Solvers over a CPU-GPU Platform. In: HPCC 2010, pp. 305–313 (2010)
17. Jung, Y., Knöpfle, C.: Real Time Rendering and Animation of Virtual Characters. IJVR 6(4), 55–66 (2007)
18. Kalbe, T., Koch, T., Goesele, M.: High-Quality Rendering of Varying Isosurfaces with Cubic Trivariate-Continuous Splines. In: ISVC 2009, pp. 596–607 (2009)
19. Kuijper, A., Heise, B.: An automatic cell segmentation method for differential interference contrast microscopy. In: ICPR 2008, pp. 1–4 (2008)

Visualization and Management of u-Contents for Ubiquitous VR*

Kiyoung Kim, Jonghyun Han, Changgu Kang, and Woontack Woo

GIST U-VR Lab., Gwangju, South Korea
{kkim,jhan,ckang,wwoo}@gist.ac.kr

Abstract. Ubiquitous Virtual Reality, where ubiquitous computing meets mixed reality, is coming to our lives based on recent developments in the two fields. In this paper, we focus on the conceptual properties of contents including definition rather than infrastructures or algorithms for Ubiquitous Virtual Reality. For this purpose, we define u-Content and its descriptor with three conceptual key properties: u-Realism, u-Intelligence, and u-Mobility. Then we address the overall scheme of the descriptor with a Context-aware Augmented Reality Toolkit for visualization and management. We also show how the proposed concept is applied in the recent applications.

Keywords: Ubiquitous VR, u-Contents, Augmented Reality, Context.

1 Introduction

M. Weiser's [14] vision of Ubiquitous Computing (ubiComp) has been being realized in our daily lives. Recently, small computing devices, such as smart phones and tablets, with higher computing powers and equipped with wireless communication capabilities have become available. Additionally, Internet access is available almost everywhere in the world. As a result, the computing paradigm has shifted towards ubiComp, in which computing resources are invisibly distributed everywhere at anytime according to the considerable developments that have been implemented in both Internet infrastructures and computing hardware. In addition to the paradigm shift, a new type of content has emerged by virtue of mashups. For instance, virtual buildings are connected to the corresponding real ones, such as SecondLife [1]. Contents generated by a mashup of geographic data with real images are used in navigation services. The contents that are mainly exploited in Virtual Space have been involved with real environments more and more.

Ubiquitous Virtual Reality (U-VR) is a new computing paradigm established on ubiComp and Virtual Reality (VR) [11]. The basic concept of U-VR is based on the seamless mapping between a virtual world and the corresponding real world. Lee et al. [11] suggested three key terminologies to be used when explaining U-VR.: context, activity and reality. A new concept and representation of contents is

* This work was supported by the Global Frontier R&D Program on <Human-centered Interaction for Coexistence> funded by the National Research Foundation of Korea grant funded by the Korean Government (MEST) (NRF-M1AXA003-20100029751).

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 352–361, 2011.

mandatory in order to support the key terminologies from a viewpoint of contents. For example, mobile Augmented Reality (AR), which overlays the informative contents onto real environments using mobile devices is a representative example of U-VR realizations. However, it is inefficient when used to represent multimedia in mobile AR environments in a conventional way because much relevant information from exterior sensors and services, such as users' profiles and locations, cannot be represented. The contents can exist not only in the virtual world like SecondLife, but also in real environments with the help of AR technologies. In this case, we need to have a hybrid description for the fusion of environmental context and ordinary contents.

A relatively small amount of work has been done to define a new concept of contents for U-VR environments. As an initial work, Oh et al. [13] introduced u-Contents, contents for U-VR. They particularly defined and explained u-Contents with a relevant scenario. Kim et al. [8, 9] reviewed the concept and discussed realization issues. Barakonyi and Schmalstieg [2] proposed intelligent agents for AR with consideration for ubiComp environments. They showed how AR agents help the given task in real applications. [2]'s work has been described from the viewpoint of an agent. For the last few years, standardization groups have researched multi-media applications, such as MPEG-7 (Multimedia Content Description Interface), MPEG-21 (Multimedia Framework), and MPEG-V (Information Exchange with Virtual Worlds), to reflect changes in the computing environment to multimedia contents (MPEG, 2009). Although the previous studies, including those involving MPEGs have dealt with some important issues for content representations and behaviors, there are still ambiguous issues in terms of relating contexts to contents in practical applications with current trends of new computing environments.

In this paper, we present a new definition of contents used in U-VR, which is an extension of Kim et al.'s work [9]. The previous work [9] did not explain practical issues in reference to u-Contents. In this work, we address properties and descriptors of u-Contents for general and practical U-VR applications. First, we review the u-Contents definition and we introduce three key properties: u- Realism, u-Intelligence, and u-Mobility, which are derived from key U-VR terminologies: reality, context, and activity. Then we reveal the relationships among the three properties. Based on these relationships, we discuss the descriptor with Context-aware Augmented Reality Toolkit that facilitates the usage of u-Contents for U-VR applications. We also present preliminary applications to show how the proposed u-Contents are used in U-VR environments.

2 u-Content Definition and Properties

2.1 Definition

We define u-Contents as multimedia contents used in U-VR environments. In particular, u-Contents realistically mediate real space, intelligently respond to user-centric contexts, and seamlessly migrate among selected entities. The terminologies we use for defining u- Contents are as follows:

- *Ubiquitous Virtual Reality* (U-VR): an emerging environment in which a collaborative wearable context-aware mediated reality is realized [11].
- *Mediated Reality*: an environment in which not only visual material can be added to augment real world experiences, but where reality may also be diminished or otherwise altered, if desired [12].
- *User-centric Context*: user-centric information among a variety of contexts in service environments that is interpreted in terms of 5W1H (who, when, where, what, how, and why) for applications [5].
- *Selected Entity*: closely related entities in one community. We follow Dey's [4] definition that the entity can be any person, place, or object.

2.2 Properties

In this section, we describe u-Contents properties in detail. u-Contents have three properties: (1) u-Realism, (2) u-Intelligence, (3) u-Mobility. u-Realism of u-Contents is necessary for the realization of mediated reality. u-Intelligence is a property which makes contents respond intelligently, while u-Mobility covers seamless migration among devices or/and spaces. Fig. 1 shows the three properties of u-Contents and their continuum. We annotate the example of contents in space. The continuum starts from existing multi-media contents and ends at the proposed u-Contents. The scale is determined by the weight of each property. We also address issues that should be resolved in the contents.

u-Realism. u-Realism is a realistic mediation that adds virtual contents into or removes real entities from real space by reflecting users' and environmental contexts. u-Realistic contents are seamlessly registered with a physical space in terms of users' senses. The visual registration overlays the virtual contents onto the real image as if the virtual contents are parts of the real one. The visual sense, in particular, is the most important aspect for seamless registration rather than tactile or auditory senses. That would be why most research in AR mainly concentrates on tracking or registration methods for the stable and accurate visual registration of contents. However, the image formation pipelines of the virtual contents and the real images are basically different so that this difference reduces the realism of the contents. Enhancement of u-Realism is achieved by reducing the differences between the augmented contents and the real scene. Klein and Murray [10] proposed a method that simulates each image formation step in order to reduce visual differences between real images and virtual contents. Besides matching different resolutions between two different contents in each space, we also need to consider the calibration and tracking techniques for seamless registration of the contents. In the continuum, 'decoupled' stage means that the contents are separated from real space. 'loosely-coupled' means that the contents are linked and registered with real entities but they can be distinguished easily. 'seamless' means that we cannot know whether the augmented one is real or not.

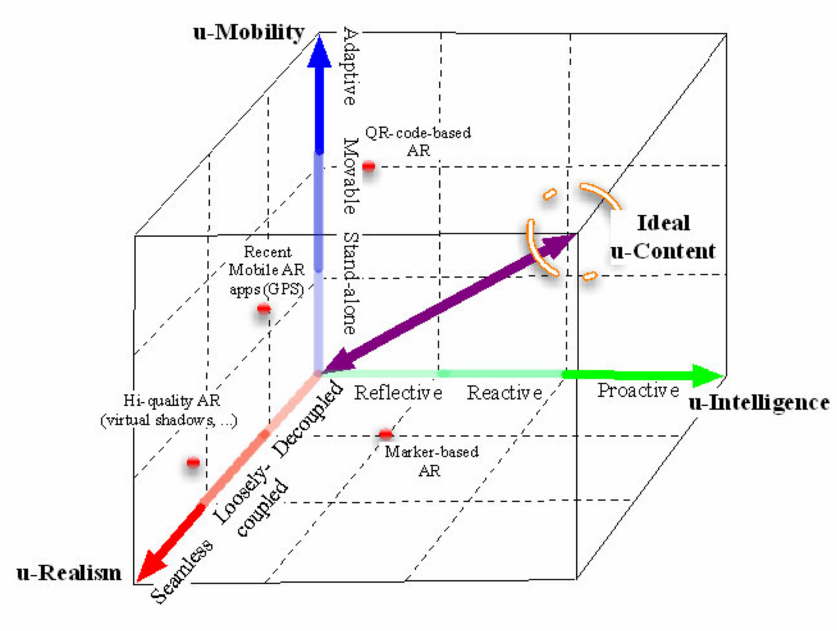


Fig. 1. u-Content continuum; each axis represents three properties, u-Realism, u-Intelligence, and u-Mobility. Each property has three keywords that can explain the level of completion. We map the existing contents or services to our continuum.

u-Intelligence. u-Intelligence is a property in which contents respond to situational information with respect to a user and adaptively change representations according to a user's explicit interaction and implicit states, such as intention, attention, and emotions [3]. The goal of u-Intelligence is to offer users good experiences by providing personalized and responsible contents in U-VR environments. To effectively support personalized experiences, u-Intelligent contents should understand the user's characteristic. Therefore, we need context-awareness to make contents reflect users' contexts. The contents also should perceive the users' interactions and respond to their interactions to allow users to interact with the intelligent contents. To make responsible contents, we should consider the method to decide content's behavior and how to express the content's responses at various situations (e.g., gestures, using image and 3D models, and sound). Then contents have memory to store past experience and the stored experience is used as the source to decide a behavior. In the continuum, 'reflective' responses happen through external factors directly. 'reactive' responses occur through a user, an object, and environments. The contents behavior is determined by processing contexts. 'proactive' responses happen through the prediction and the forecast. The contents memorize its experience and predict situation through the experience.

u-Mobility. u-Mobility is a property that enables u-Contents to move among selected entities. It includes contents that move from one environment to others and to selected entities. u-Contents can be transmitted and translated in accordance with characteristics of u-Mobility. There are three key issues to be considered in order to achieve the u-Mobility property. The first one is content transmission, for example, minimum requirements of network bandwidth and supported protocols. It is necessary to consider the quality of contents and to support real-time visualization. After the transmission, context-awareness, especially environmental context information, is required because u-Contents have to be adapted to the new environment. When we consider u-Contents as multi-media contents that exist in real, virtual, and mixed environments, u-Mobility is necessary to allow the contents to migrate between those environments. The third issue of u-Mobility relates to privacy when contents are shared. When contents move to other users' entities, disclosure and privacy control have to be considered. In order to determine how to share u-Contents with others, personal and social context have to be considered. The social context is realized by user social network among users and properties of u-Contents. In the continuum, 'stand-alone' stage means that the contents can be viewed through homogeneous devices. 'movable' content can be viewed through heterogeneous devices by migration. The last 'adaptive' stage means that we can view the contents in any devices according to dynamic situations through changing properties of content in accordance with target environments or devices.

3 u-Content Descriptor for Visualization and Management

It is important to formulate a context in a consistent way. To achieve consistent context formulation, we have adopted Jang and Woo's [6] Unified Context-aware Application Model (UCAM) in the 5W1H (who, when, where, what, how, and why) form. The proposed descriptor is designed to include not only the general meta-data of a context, but also descriptions of each u-Contents property in the 5W1H form. Representations of contents in the 5W1H form allow U-VR applications to exploit user contexts and environmental contexts in their contents rendering process in two ways, directly and indirectly.

3.1 Descriptor Format

The u-Contents descriptor is a header of u-Contents in a meta-data form. The descriptor defines the characteristics of multimedia contents that are being followed. Extensible Markup Language (XML) is adapted for formatting data. Note that u-Contents descriptors should be written from a contents viewpoint, not an applications viewpoint. Table 1 shows the u-Contents descriptor that classifies the u-Contents into each 5W1H field.

Table 1. u-Content descriptor format in 5W1H

Field	Description
Who	The author (owner) of the contents
When	The last creation or modification time
Where	The position of the contents in a local or a global coordinate
What	Parameters for u-Contents properties: [u-Realism, u-Intelligence, u-Mobility]; Refer Table 2 more in detail.
How	The information about types of authoring tools or context-aware toolkits involved with all fields
Why	The last inferred results obtained from a certain context-aware toolkit

Table 2. Description of 'What' field

Property	Static attributes	Dynamic attributes
u-Realism	**Quantitative information of contents body**: (e.g.) size or dimension of data, total mesh numbers of a 3D graphics model, etc. **5 senses representation:** (e.g.) data representation considering sight, hearing, touch, smell and taste.	**Context history of contents body**: (e.g.) pre-contexts are kept for other applications **Environmental information:** (e.g.) light source positions, temperature, intensity of illumination, etc.
u-Intelligence	**Quantitative information:** (e.g.) the list of possible responses, etc. **Additional information for intelligent responses**: (e.g.) the list of autonomy control rules based on content-specific information.	**User-specific information:** (e.g.) user-specific logics **Environmental information:** (e.g.) a user location, profile, etc.
u-Mobility	**Privilege of u-Contents**: (e.g.) an ownership of u-Contents, privilege of entities sharing the ownership, the level of disclosure and permitted modification. **Requirement for movement**: (e.g.) properties of a target entity: minimum bandwidth of network, supported protocols, etc.	**Community information**: community meta-data (e.g.) members, the goal of the community and relations among the entities. **Movement information**: (e.g.) the movement information is kept for understanding usage of the u-Contents.

The descriptor can be compatible with contents used in common authoring tools except for the 'What' field. The 'What' field, where the properties of u-Contents are mainly described, has two types of attributes, static and dynamic. While the static attributes are not changeable, the dynamic attributes can be overridden whenever the u-Contents are passed to context-aware modules for processing. We rely on these simple criteria of types to classify and to group characteristics of u-Contents. Detailed descriptions of the 'What' field are shown in Table 2.

The proposed descriptor can be extended to represent the meta-data of existing content formats because it includes the information in existing 3D models, videos, images, and so on. There is less possibility of experiencing compatibility problems when using the proposed descriptor in conventional contents and context-aware applications. We found similar approaches in MPEGs. MPEG-7 has a descriptor for audio-visual contents and MPEG-21 has a digital item that is the fundamental unit (meta-data) of distribution and transaction within the MPEG-21. We expect that the combination can be implemented within our descriptor by rearranging detail components.

3.2 Descriptors in a Context-Aware Augmented Reality Toolkit

We display the framework, Context-aware Augmented Reality Toolkit (CART), to validate the usages of the proposed descriptor. The main goal of the CART is to facilitate the usage of various contexts and contents for U-VR applications. Basically, the CART encodes and decodes u-Contents descriptors for U-VR applications. Hence, the CART allows developers to easily fuse AR and context-aware technologies for U-VR applications. The CART consists of a context-aware module and an augmentation- enabled module. Fig. 2 shows how u-Contents are used as one input of CART. When the u-Contents are delivered to CART, the u-Contents descriptor (D) is separated from its body. Then, the descriptor is sent to UCAM to be integrated with all contexts from other services or sensors in the environments. The context-aware module has functions to communicate, integrate, and manage contexts represented in the 5W1H form. UCAM supports three core functionalities: communicator, context integrator, and context manager. The communicator gathers all raw signals from sensors via networks and passes them as preliminary context (*p-context*). In the context integrator, the *p-context* is transformed to the integrated context (*i-context*) by fusing available contextual information. The context manager finalizes the given *i-context* with predefined context and packages them as final context (*f-context*). The *f-context* is transmitted by the communicator and shared with other U-VR applications. The u-Contents descriptor also goes through the same path, and contextual parts are selectively overwritten during the process.

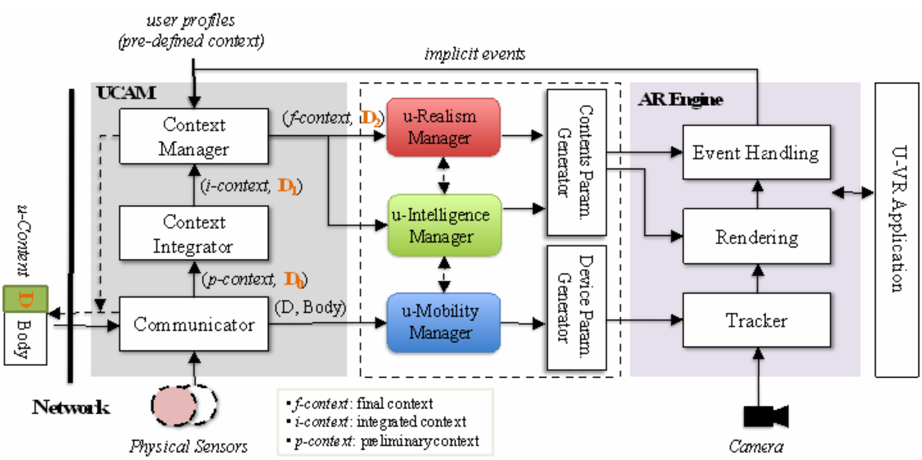


Fig. 2. Flow of u-Contents in the Context-aware Augmented Reality Toolkit

The context cannot be sent to the augmentation module directly because the scales of parameters in the context module are different from the scales in the augmentation module. Consequently, we need another manager in order to change parameters into the scale of AR environments. u-Reolism, u-Intelligence, and u-Mobility managers reverse the scales of parameters. The u-Realism manager extracts the defined information from the f-context and transforms it into rendering parameters. While the u-Intelligence manager determines responses, the u-Mobility manager directly receives contexts from sensors because it is related to low level processing, such as parameters for tracker modules.

4 Examples

In this section, we introduce simple examples and discuss the contents from the viewpoint of the proposed concept. Although the existing examples do not explain the u-Contents concept in all aspects, they can help us understand the u-Contents and properties.

4.1 Virtual Character in the U-VR Room

Fig. 3 shows the snapshots of the corresponding places in real, virtual, and augmented space. If we think this as a preliminary U-VR space, the bluebird in the room could be u-Content. We designed the 3D bluebird model that acts like a virtual pet as the avatar. Basically, the bluebird reacts against two outer contexts, a user's location and environmental sound information. The bluebird has basic predefined behaviors, such as 'fly,' 'walk,' and 'run'. Our assumption is that the user can see the bluebird via a Head-Mounted Display (HMD) or personal mobile devices. When a user enters the room, the bluebird is augmented onto the most silent place, the place furthest from the user. After the user sits on a sofa, the bluebird tries to move toward the user. When a sound such as clapping, is generated, the bluebird returns to the silent place again.

4.2 3D Content Interacting with a Real Object

u-Intelligent contents should respond to users or objects including virtual and real objects, according to situational information. Fig. 4 shows 3D content responses according to situational information. In Fig. 4, there is a 3D augmented character that has predefined behaviors. In Fig. 4(a) and Fig. 4(b), the 3D character perceives a relative position and direction (velocity) of the real toy cart, and determines to push or pull the toy cart. In Fig. 4(c), the 3D character is collided to the toy cart and falls down, as it perceives direction and speed of the toy cart. Because the 3D character considers content's responses to a real object's status, this example is not a perfect u-intelligent content. It should consider how to be aware of user's characteristics and respond adaptively in order to achieve u-intelligence.

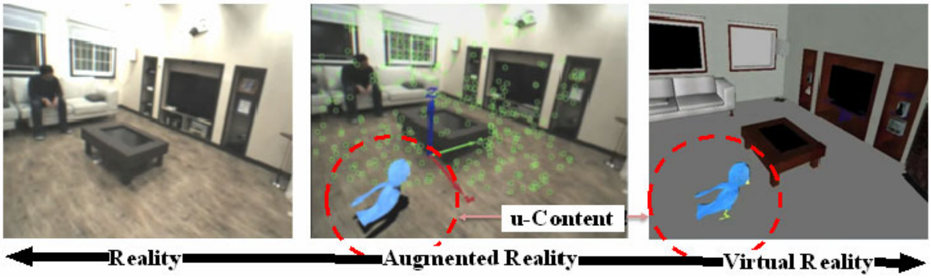


Fig. 3. Virtual bluebird in the U-VR room. The environment has three modes; real, augmented, and virtual reality.

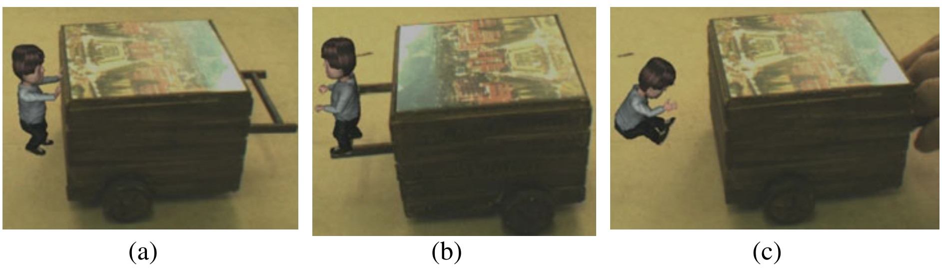


Fig. 4. 3D character responses according to geometrical and conditional information between 3D character and a toy cart: (a) push, (b) pull, and (c) fall down

5 Conclusion and Future Work

We presented u-Contents concept with the visualization and the management schemes. And also we introduced recent on-going works and their relationship within the u-Contents continuum. Two examples in Section 4 showed the preliminary works to explain our concept. The bluebird example can be mapped 'loosely-coupled', 'movable', and 'reactive' stages. The example dealing with content interaction can be mapped 'loosely-coupled', 'stand-alone', and 'reactive' stages. As one of consecutive works, the standardization activity is initiated [7] to represent mixed reality contents. However, we expect that there will be many practical obstacles to be resolved. One of them is "Contents Ecosystem". In the future, u-Contents and their **Ecosystem** should be discussed together to verify the availability of the proposed contents concept in daily life.

References

1. Second life (2009), `http://secondlife.com`
2. Barakonyi, I., Schmalstieg, D.: Ubiquitous animated agents for augmented reality. In: Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2006, pp. 145–154 (2006)
3. Cho, S.B., Kim, K.J., Hwang, K.S., Song, I.J.: Anidiary: Daily cartoon-style diary exploits bayesian networks. IEEE Pervasive Computing 6, 66–75 (2007)

4. Dey, A.K.: Understanding and using context. Personal Ubiquitous Comput. 5, 4–7 (2001)
5. Hong, D., Schmidtke, H.R., Woo, W.: Linking context modelling and contextual reasoning. In: 4th International Workshop on Modeling and Reasoning in Context (MRC), pp. 37–48 (July 2007)
6. Jang, S., Woo, W.: Ubi-ucam: a unified context-aware application model. In: Proceedings of the 4th Inernational and Interdisciplinary Conference on Modeling and using Context, CONTEXT 2003, pp. 178–189 (2003)
7. Kim, H., Yoon, H., Choi, A., Baek, W., Lee, I., Kim, D., Woo, W.: Data markup representation for mixed reality contents. In: International AR Standards Meeting (2011)
8. Kim, K., Hong, D., Lee, Y., Woo, W.: Realization of u-Contents: u-Realism, u- Mobility and u-Intelligence. In: International Symposium on Ubiquitous Virtual Reality, pp. 3–4 (July 2007)
9. Kim, K., Oh, S., Han, J., Woo, W.: u-Contents: Describing contents in an emerging ubiquitous virtual reality. In: International Workshop on Ubiquitous Virtual Reality, pp. 9–12 (February 2009)
10. Klein, G., Murray, D.: Compositing for small cameras. In: Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR 2008, pp. 57–60 (2008)
11. Lee, Y., Oh, S., Shin, C., Woo, W.: Recent trends in ubiquitous virtual reality. In: International Symposium on Ubiquitous Virtual Reality, pp. 33–36 (July 2008)
12. Mann, S.: An historical account of the 'wearcomp' and 'wearcam' inventions developed for applications in 'personal imaging'. In: ISWC 1997 (1997)
13. Oh, S., Lee, W., Park, Y., Woo, W.: u-Contents: New kinds of realistic contents in ubiquitous smart space. In: International Symposium on Ubiquitous Virtual Reality, pp. 13–16 (July 2006)
14. Weiser, M.: The computer for the 21st century. SIGMOBILE Mob. Comput. Commun. Rev. 3, 3–11 (1999)

Semi Autonomous Camera Control in Dynamic Virtual Environments

Marcel Klomann and Jan-Torsten Milde

Fulda University of Applied Sciences, Department of Computer Science
marcel.klomann@informatik.hs-fulda.de, milde@hs-fulda.de

Abstract. We present a system for controlling the camera movement in an interactive dynamic virtual world. The camera control is scripted in a specialized scripting language. A portable script interpreter has been implemented, allowing to run the scripts both on standard PCs and XBOX 360 systems.

Keywords: Camera control, script interpreter, PC and XBOX.

1 Introduction

Interacting with semi autonomous agents in dynamic virtual environments has been in a central research topic over the last decades. Agent control has become more reactive and intelligent, making the agents' actions more realistic and user friendly. Camera control, on the other hand, has not been in focus of research so much. In most virtual environments the camera is realized as a simple follow up camera or it takes on the first person perspective of the currently active agent. This simple approach has a strong negative impact on the way interactive stories are constructed in virtual environments.

Therefore, in order to make the human computer interaction believable, an adaptive and intelligent camera control is essential. Outside the realm of virtual environments a team consisting of director, cameraman and cutter are responsible for the selection of angle, focus and movement of the camera. In our approach we like to transmit this idea of directing the camera into the field of virtual environments.

As a test bed we developed an XNA-based game engine for the XBOX 360 and standard PCs (see 2). The game engine allows us to display complex 3D virtual environments. Animated agents populate the world, assets can be placed, the agents are able to manipulate the assets and can interact and communicate with other agents (see 5).

The central part of the game engine is a portable script language interpreter targeting the intelligent and adaptive camera control. An easy to learn C-like syntax has been defined for the script language. The script interpreter and the game loop are seamlessly integrated, making camera control simple, yet powerful.

Scripts can be attached to any kind of agents within the virtual environment. This also includes cameras as a specific kind of agent. As such, the script language, could also be used to control the agents, which is a nice side effect of the development. It should also be noted that up till now, no functional script interpreter existed for the XBOX 360.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 362–369, 2011.

The camera control language abstracts away from the underlying mathematical concepts. Instead the commands mimic the language used in film making. The script language includes control statements making it possible to react to dynamic state changes in the virtual environment.

All parameters of the camera can be controlled over simple to use commands. The camera control parameters include position, focus, zoom, angle, field of view, animated movement, follow up behavior, path following and even visual special effects, using pixel shaders for image post processing.

The following script defines two positions in the virtual environment. The camera is focusing an agent and then continuously moves from position p1 to position p2 without losing focus on the agent:

```
#import "vector.sfx";

Vector3 p1 = Vector3{10.0, 100.0, -300.0};
Vector3 p2 = Vector3{10.0, 100.0, -300.0};

Focus(agent);

// Move from one point to another
// in an infinite loop.
while(true)
{
 Wait(MoveToPoint(p2, 10.0));
 Wait(MoveToPoint(p1, 10.0));
}
```

Listing 1: Example script with moving camera

Even as the agents moves through the world, the camera will not lose focus on him. The camera's movement is controlled independently of the agent's actions. It will continue in parallel as long as the director is not giving new commands, aka attaching a new script to the camera. Attaching and detaching scripts can also be accomplished with a script command. As such, a the behavior of a camera can change over time. Finally a message based communications system has been implemented, that enables the exchange of data between multiple scripts.

I our experiments the described approach has shown to be very versatile. The camera control is clearly separated from the agent control. The compact syntax of the camera control language makes it simple to define complex camera movements, which are both controlled, yet reactive. The scripts tend to be small in size, and are easy to program and understand. The script language supports a more film oriented style of camera control, thus enhancing the quality of story telling in virtual environments.

2 Game Engine Architecture

Fundamentals: it's not possible to develop a whole game engine during a single project year (lack of time and students do not meet the necessary requirements). Nevertheless the students still were able to develop playable games, having their focus of development on the visual parts of the game (rendering and animation). Using the Microsoft XNA Framework, the games were developed in C# and runnable on PC and Microsoft XBox 360.

Based on the results of the course, a new game engine was developed. The goals of the game engine are: Previsualization, simple definition of virtual worlds, realistic lighting and rendering, simple development of 3D games with mixed environments.

The developed script language, as a central part of the game engine, aims for a simple camera control. Conceptual, the camera control orientates itself by the vocabular of film creation, which means it abstracts from the underlying linear algebra and does not require knowledge of it.

The goal of the game engine architecture is to provide simple to use components and base classes the students can use to achieve fast results (see 3).

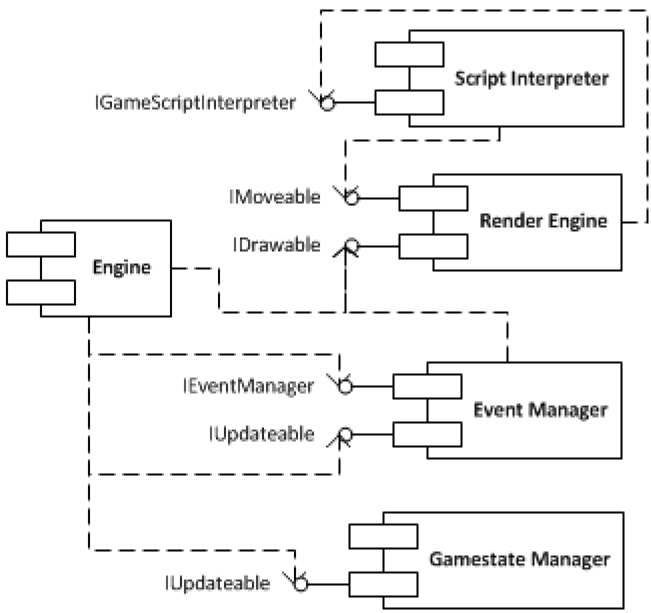


Fig. 1. Components of the game engine architecture

The script interpreters main task is to move objects in the virtual world. Therefore moveable objects implement the interface IMoveable or extend the base class MoveableObject, which provides basic movement functionality for objects. Both make objects controllable by the developed script interpreter. The MoveableObject class provides the fundamentals for the implemented camera class. The lighting model consists of several local light sources, which are spot lights and point lights.

Those light sources are also capable of drawing a shadow map and casting realistic shadows using the created shadow map. There is also a directional light source, which does not cast shadows. Spot lights and point lights make use of the provided classes for moveable objects so light sources are controllable via scripts as well, just like the camera class.

3 Script Language

Currently, there is no functional script language supporting the XNA Framework and especially not a script language supporting XBox 360. Therefore a script interpreter was developed, supporting game development on PC and XBox 360.

The ANTLR (see 1 & 5) class library for C# was used to generate the parser classes from a formal grammar. The main part of the development was the integration of the script interpreter into the game loop. On the one hand, the scripts must execute concurrently and seamlessly to the game loop. On the other hand, the developers using the script interpreters shall not bother with the integration into the game loop or managing the scripts seamless execution. To meet those requirements, the scripts must be able to stop execution at a specific point and continue later. That means each script shall have the opportunity to execute on an update cycle of the game loop, stop and continue on the next update cycle. To achieve this, a seperate thread for each running script is used, which makes it easy to make scripts wait for a specific event to occur using thread synchronization mechanisms.

This way, scripts can synchronize with the game loop by using an event-based mechanism to start events that take several frames to complete, wait for their completion by the game loop and resume afterwards.

3.1 Language Design

For a better learnability the syntax of the script language orientates by existing languages. A simpler version of the C syntax was chosen to provide an easy syntax that is also well known by many developers. So the language design is simple and imperative. Functions are chosen to provide a way to interact with the game engine. The mentioned synchronization with the game loop also makes use of functions to create and wait for specific events.

The script language also provides primitive data types like Integer, Char, Float, Boolean and String. In addition to that, the language also supports the definition of custom structs. This can be used to work with more complex data types like vectors or even matrices.

To support a modular script code, import statements are supported and to keep the scripts well documented, single line and multiline comments are also supported.

3.2 Script Interpreter

As already mentioned, ANTLR was used to generate a parser from a formal grammar. The developed ANTLR grammar was used to generate a lexer, parser and tree parser. In a first pass, the lexer makes a lexical analysis of the script to find "words" (so called tokens) which are then used in a second pass by the parser to recognize the grammatical structures of the sentences (grammar rules). The parser generates an Abstract Syntax Tree (AST) which is later used for an efficient execution of the script. The script execution happens in a third pass. The generated tree parser traverses the generated AST, recognizes commands (e.g. function calls) and passes them to the actual interpreter, which will execute the command then.

The interpreter itself works as a one-pass system, which means that function and struct definitions are recognized in the execution pass. This restricts those definitions to appear in the source code before they are used the first time.

Script execution time is only a few milliseconds, which is required by a real-time application like a game engine. However, on the first script execution, the Just-in-Time-Compiler needs to compile the required sources first, so that script execution takes a lot longer then. But this has no negative impact on practical applications.

A big advantage of this script interpreter is the support for Microsoft XBox 360. It's fully compatibel with both systems, PC and XBox 360, while other interpreters (e.g. LuaInterface, see 4) are limited to PCs. But performance differences caused by the different hardware of the XBox 360 must be taken into account, too.

3.3 Scripted Camera Control

The current application of the script interpreter is the camera control. Using abstract commands that orientate by camera directions for a camera man (panning, zooming and movement), a simple camera control arises that doesn't require knowledge of linear algebra or 3D programming. Therefore, the commands only interact with the event mechanism by creating events and waiting for their completion. The created events are completed by the game loop of the game engine. When an event is completed, the game loop notifies the script so it can resume.

This way, complex camera movements can be accomplished with only a few commands while details of the implementation, like matrix multiplications, stay hidden.

The following is a sample script that moves a camera along a specified path and also keeps focus on a point in space:

```
#import "vector.sfx";

Vector3 target = Vector3{100.0, 50.0, 100.0};
Focus(target);

// Define a new path for the camera.
// Each key point will be
// approximated by a mathematical curve at the specified
// point in time. The used mathematical curves provide a
// smooth approximation and interpolation
// which results in a
// smooth movement of the camera.

int pathID = NewPath();
AddKey(pathID, 0.0, Vector3{-5.0, 50.0, 165.0});
```

```
AddKey(pathID, 5.0, Vector3{50.0, 70.0, 100.0});
AddKey(pathID, 10.0, Vector3{150.0, 90.0, -200.0});
AddKey(pathID, 15.0, Vector3{200.0, 110.0, -400.0});
AddKey(pathID, 20.0, Vector3{-40.0, 70.0, -400.0});
AddKey(pathID, 27.0, Vector3{-5.0, 50.0, 165.0});

// Tell the camera to move along the
// previously defined path,
// which results in a unique ID that can be
// used to wait for the
// completion of  the task. The move takes 27 seconds.

int waitID = Move(waitID);
Wait(waitID);
```

Listing 2: Camera movement with fixed eye point

It takes just a few lines of code to let the camera move along a long path. The synchronization with the game loop is done by the call of the function "wait", which waits for the completion of the specified event.

Die beteiligten Komponenten besitzen folgende Struktur:

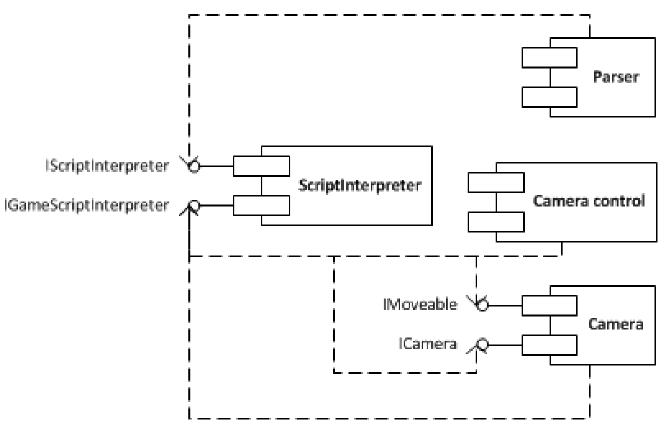


Fig. 2. Components of the scripted camera control

3.4 Extending the Language

The language was not only designed for the use in a camera control, but to be used in almost every application in a game engine. It's already possible to control every moveable object that extends the class "MoveableObject". Furthermore it is possible to extend the script interpreter by adding more script functions that provide an interface between scripts and game engine. Methods programmed in C# can be added as script functions using delegates and will be called in case the corresponding function is called inside a script. Therefore, scripts can be used to control Non-player characters (NPCs), in-game events or the whole game (A.I.).

In addition the script interpreter is not limited to applications in game engines, but can also be used in ordinary applications.

4 Future Work

Using multiple threads to achieve a seamless concurrent execution of scripts and game engine brings major performance issues in case that many scripts are executed at once. This is primarily caused by frequent context switches and a default stack size of one megabyte per thread, which is a huge memory waste on systems like the XBox 360.

For those reasons, another solution to achieve a seamless execution of scripts, that does not depend on a separate thread per executed script, must be found.

This requires that scripts can be paused at any time and continued later. To implement this in a single thread, the execution context of a script must be backed up on each interruption and restored on continuation. Considering the call stack of the CLR is write protected, this is not possible. This also means, the script interpreter is not capable of achieving a seamless execution without using multiple threads or without major restructures. So a suitable script interpreter requires a VM maintaining his own call stack. This also excludes parsing a hierarchical structure, like the AST, to interpret a script, because parsing such a structure still requires parts of the execution context, that are important for script interpretation, to be stored on the call stack of the CLR. Therefore a more suitable script structure must be provided in order to achieve a linear script execution, which does not store essential information on the call stack of the CLR. To produce a linear intermediate language, another parsing pass must be introduced. The intermediate language is comparable to assembler and instructions are not linked in any way. This enables the script interpreters VM to execute the scripts using a program counter that keeps track of the last executed instruction, making it possible to pause the script and continue at any point in time. This development shall make the system more flexible and scalable to use it in every imaginable scenario not limited by the number of concurrent scripts.

5 Conclusions

The system runs stable on both platforms, PC and XBox 360. The script language can be used to implement flexible camera controls. The system is also extensible, which makes it suitable for many other applications within game engines or in ordinary applications.

References

1. ANTLR, `http://www.antlr.org`
2. Carter, C.: Microsoft XNA Game Studio unleashed. Sams (2009)
3. Gregory, J.: Game Engine Architecture. Transatlantic Publishers (2009)
4. Ierusalimschy, R.: Programming in Lua (2006), `http://www.lua.org/`
5. Klomann, M.: Skriptgesteuerte Kameraführung in vituellen Umgebungen. Bachelorthesis Hochschule Fulda. Fachbereich Angewandte Informatik (2010)
6. Parr, T.: The Definite ANTLR Reference. Building Domain Specific Languages. Pragmatic Programmers (2007)

Panoramic Image-Based Navigation for Smart-Phone in Indoor Environment

Van Vinh Nguyen, Jin Guk Kim, and Jong Weon Lee

Sejong University, Seoul, Korea
vinhnv2@yahoo.com, jingoogi@hotmail.com, jwlee@sejong.ac.kr

Abstract. In this paper, we propose a vision-based indoor navigation system for a smart-phone. The proposed system is designed to help a user traveling around an indoor environment to determine his current position and to give him the direction toward a chosen destination. For sensing user's position and orientation, the system utilizes panoramic images, which are pre-captured the environment and then processed to create a database. For matching images captured from user's smart-phone with the database, we use SURF[1], a robust detector and descriptor. Besides, to minimize responding time, the system employs client-server architecture in which a server module is mainly in charge of time consuming processes. Also, a tracking mechanism is applied to reduce matching time on the server. The experimental results show that the system can work well on a smart-phone in interactive time.

Keywords: Indoor navigation, panorama tracking, augmented reality.

1 Introduction

Generally, navigation in unknown environment considers Global Positioning System (GPS) to estimate positions. Many GPS-based navigation systems running on robot or vehicles are available and work effectively. Recently, with the growing of new generation of smart-phones that are equipped with GPS sensor, navigation on a smart-phone becomes realistic. However, a GPS-based navigation method only works well in outdoor environment. Since GPS signals are not available for indoor environment and other sensors such as WiFi-based estimation [2] and RFID-based tracking [3] are not robust yet, in this paper we focus on vision-based navigation method that is normally simpler, cheaper and easier to deploy broadly than other solutions.

Vision-based localization methods are normally based on the analysis of captured images to get information of current position. In our proposed system, indoor localization is based on the comparison of runtime captured images with the database of reference images, which are pre-captured from the targeting indoor environment. Each image can be described as a set of natural features, thus the comparison of two images becomes feature matching problem. To deal with this problem, we apply a SURF feature [1] which has been found to be highly distinctive and repeatable. However, SURF-based matching shows low speed when running on a smart-phone due to limited resources of the device, thus we employ client-server architecture to

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 370–376, 2011.

reduce system's responding time. Because reference images in the database should cover entirely the targeting environment for efficient localizing, we rely on pre-captured panoramic images to reduce the number of required reference images.

The remainder of this paper is structured as follows. Section 2 surveys related works in vision-based mobile phone localization and natural feature descriptors. Section 3 describes our approach in detail. Experiment results and discussions are presented in Section 4. We conclude with a discussion of limitations and future work in Section 5.

2 Related Works

The AR-PDA project [5] develops an image-based object recognition and tracking method for AR applications that requires the 3D-model of an object to be matched. A fully indoor navigation system targeting on camera phone has been proposed in [6], which is based on floor corners and SIFT features. This work relies on client-server architecture to reduce response time, but the system still suffers from late response that is about 10 seconds. S. Saito et al. takes a different approach. They present their indoor positioning system [7] based on fiducial markers that are pre-designed for containing position information. Using markers for determining user's position is easy way to deploy, but it locks a user to fixed positions for sensing current position. In our proposed approach, we try to make a user freely move around environment while getting navigation guides at anywhere he is staying. To enable this feature, we apply natural feature-based matching method for our system.

For comparing two different images, many approaches have been proposed. Point-based approach uses interest point detectors and local descriptors to localize and describe image features. SIFT (scale invariant feature transform) method [8] is well-known for its robustness. The SIFT feature descriptor is invariant to scale, orientation, affine distortion and partially invariant to illumination changes. A research mentioned in [9] indicates that the SIFT-based descriptors outperform other local descriptors on both textured and structured scenes. The SURF descriptor [1] is partly inspired by the SIFT descriptor and has similar performance to SIFT, while it is much faster than SIFT. Considering the system performance, we choose SURF as our main detector and descriptor.

3 Proposed System

3.1 System Overview

Considering the overall performance, the proposed system is designed in client-server architecture. It consists of a client module running on smart-phone(s) connected to a server via a wireless network. The system's architecture is demonstrated in Fig. 1.

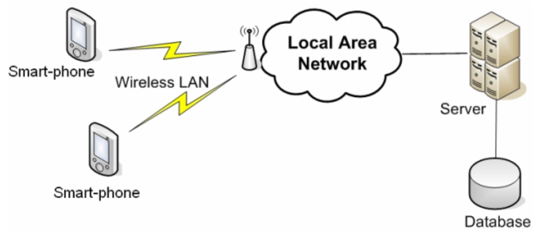


Fig. 1. System architecture diagram

The client module uses smart-phone's camera for capturing images of the real world and sends those images to the server. When the server receives an image, some filters are applied to the image to remove noises. After that, the received image is compared with ones stored in the database. The comparison of images is based on SURF features [1] that are detected in the image. If any match is found, the server estimates current user's position and orientation and they are sent back to the smart-phone. Based on the estimated position and orientation, the client module will display a direction to a chosen destination. Fig.2 shows the system's processes and their relations.

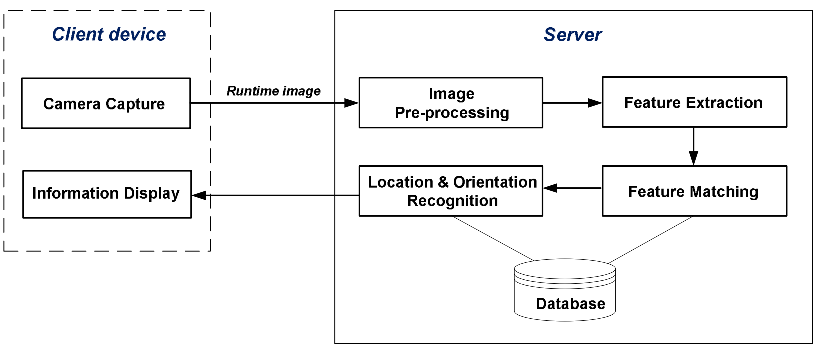


Fig. 2. System's workflow diagram

3.2 Database Creation

Because our position recognition approach is based on image matching, we need to create a database of features of reference images. Features on the DB are compared with features of smart-phone's images during the runtime phase. A large set of reference images should be taken so that they cover the targeting environment. For minimizing the number of required images, we capture panoramic images that are taken in the environment. Specifically, creating the database has following steps:

- *Step 1*. We setup a 2D reference coordinate system in the indoor environment and then the coordinates of featured places (i.e. rooms in a building) are roughly estimated.

- *Step 2*. For each featured place, we create several 360^o panoramic images. Panoramic images are easily created by using some available applications such as panorama tools using a stitching method. In our work, we use wide-angle camera lens for capturing environment and then use a stitching tool to create panoramas. Besides, the number of panoramic images in each featured place has to be minimized to reduce the matching time while maintaining the required resolution of position estimation. Practically, we divide each place into 4 m x 4 m cells and take one panoramic image in the center of each cell. The position and orientation are recorded with each panoramic image. The orientation of panoramic image means the angle between the view's center and Y-axis of the reference coordinate system.
- *Step 3*. We divide horizontally each panoramic image into four sub-images. That means one sub-image covers 90^o of the view and it is easy to calculate the orientation of each sub-image based on the orientation of the original panoramic image. Fig. 3 shows one set of sub-images of a panoramic image.
- *Step 4*. We run SURF detector and descriptor on each sub-image and then store the descriptors with corresponding positions and orientation information into the database.

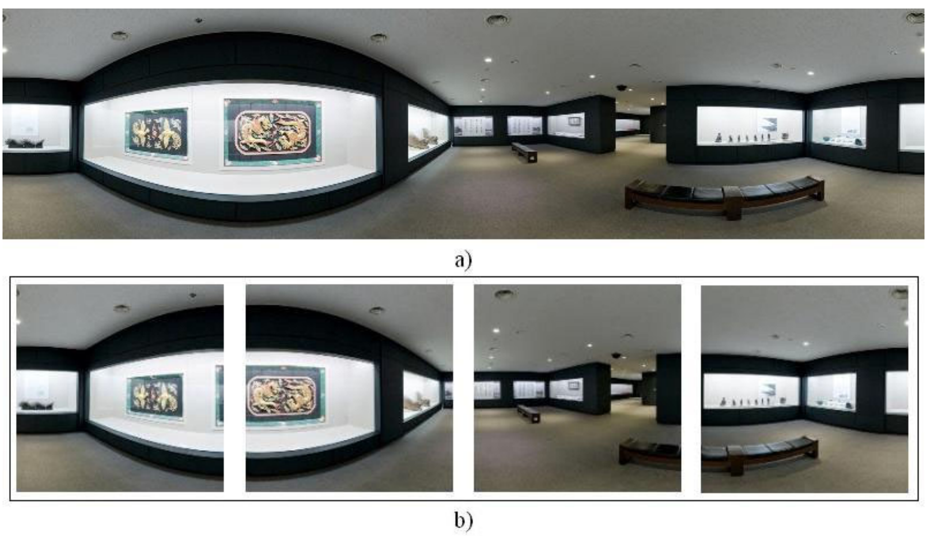


Fig. 3. a) 360^o panoramic image b) Four sub-images created from the panorama

3.3 Position and Orientation Sensing

An image sent from a smart-phone to a server is processed using SURF. Generated SURF descriptors are compared to ones in the database. If any match is found, the corresponding position and orientation of the matched sub-image in the database are considered as the position and orientation of the smart-phone. However, the captured image from the smart-phone is usually smaller than reference sub-images and these images are not often aligned. Therefore, we apply homography to revise the real orientation of the smart-phone: we use RANSAC[4] to estimate the central point and four corners of the matched reference sub-image in the captured image's frame. Having that information, we can calculate the error of the orientation using the equation 1. After estimating the current coordinates and the real orientation of the smart-phone, it is straightforward to calculate the direction from the current position to the chosen destination since any chosen destination's coordinates is known.

$$\Delta\beta = \frac{\Delta x * 90}{Ws} \tag{1}$$

Ws: Width of reference image in captured image's frame.
Δx: Difference of 2 central points' x-coordinate in captured image's frame.

To reduce matching time, not all of the reference images in the database are used to compare with smart-phone's captured image. Each captured image is firstly compared to reference sub-images that are closely located to previous position of the smart-phone. Assuming slow user's motion, this mechanism can avoids comparing a captured image to all reference images, which is significant time consuming process.

4 Experimental Results

We have implemented a test version of the proposed navigation system for performance evaluation. The client module is deployed on an iPhone-3GS whereas the server module is running on a PC-2.4GHz. The tested indoor environment is a museum and a database of features is created from 11 panoramic images. The system is then tested in the real museum with a following scenario:

Fig. 4. An example of server-side operation. Left image is one received from iPhone. Right one shows the direction to the target location.

"A user travels around the museum with an iPhone. When he needs to move to another place, he uses the iPhone to capture images (or objects) near his current position. Based on the captured image, the navigation system running on the iPhone will show a virtual arrow indicating the direction to the destination. Moreover, an estimated distance to the destination is also displayed".

The testing results show that the system can guide a user successfully to a destination in an interactive time by displaying a correct direction and an approximate distance. For each image received from an iPhone client, the server consumes around 0.7 second to determine position and orientation of the iPhone. Considering time for transferring information from the server to the phone, the average response time of the proposed system is around 1 second.

5 Conclusion

We have proposed a navigation system for mobile phone in indoor environment. The proposed system exploits the characteristics of panoramic images, which cover 360^O of surrounding scenes, to create a database of reference images. Smart-phone's positions and orientations are estimated based on positions and orientations of matched panoramas. Besides, to improve the performance of the system, the system employs a client-server architecture that outsources all time-consuming tasks to a server. Also, a tracking mechanism is applied to reduce matching time on the server. Therefore, the system achieves reasonable responding time as proven in the experimental evaluation. The limitation of the system is the difference between estimated smart-phone's position and real position. Currently, the estimated position is same as the nearest (or the most similar) panorama's position. For any applications which require more precise position of a smart-phone, we need to improve the position estimation method. This problem is addressed as our future work.

Acknowledgments. This research is supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA), under the Culture Technology(CT) Research & Development Program 2010.

References

1. Bay, H., Tuyelaars, T., Van Gool, L.: Speed-up Robust Features. In: ECCV, pp. 404–417 (2006)
2. Gezici, S.: A Survey on Wireless Position Estimation. Wireless Personal Communications: An International Journal 44(3), 263–282 (2008)
3. Bouet, M., dos Santos, A.L.: RFID tags: Positioning principles and localization techniques. In: Proceedings of IFIP Wireless Days - 2nd International Home Networking Conference, IHN 2008 (November 2008)
4. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm. of the ACM 24(6), 381–395 (1981)

5. Gausemeier, J., Frund, J., Matysczok, C., Bruederlin, B., Beier, D.: Development of a real time image based object recognition method for mobile AR-devices. In: Afrigraph, pp. 133–139. ACM, New York (2003)
6. Hile, H., Borriello, G.: Information overlay for camera phones in indoor environments. In: Hightower, J., Schiele, B., Strang, T. (eds.) LoCA 2007. LNCS, vol. 4718, pp. 68–84. Springer, Heidelberg (2007)
7. Saito, S., Hiyama, A., Tanikawa, T., Hirose, M.: Indoor Marker-based Localization Using Coded Seamless Pattern for Interior Decoration. In: Virtual Reality Conference (2007)
8. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision (2), 91–110 (2004)
9. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence (2005)

Foundation of a New Digital Ecosystem for u-Content: Needs, Definition, and Design

Yoosoo Oh[1], Sébastien Duval[1], Sehwan Kim[2], Hyoseok Yoon[3], Taejin Ha[3], and Woontack Woo[3]

[1] GIST CTI, 261 Cheomdan-gwagiro (Oryong-dong), Buk-gu
Gwangju, 500-712 Republic of Korea
[2] WorldViz, 614 Santa Barbara St, Santa Barbara, CA 93101, USA
[3] GIST U-VR Lab., 261 Cheomdan-gwagiro (Oryong-dong), Buk-gu
Gwangju, 500-712 Republic of Korea
{yoh,sduval}@gist.ac.kr, kim@worldviz.com,
{hyoon,tha,wwoo}@gist.ac.kr

Abstract. In this paper, we analyze and classify digital ecosystems to demonstrate the need for a new digital ecosystem, oriented towards contents for ubiquitous virtual reality (U-VR), and to identify appropriate designs. First, we survey the digital ecosystems, explore their differences, identify unmet challenges, and consider their appropriateness for emerging services tightly linking real and virtual (i.e. digital) spaces. Second, we define a new type of content ecosystem (u-Content ecosystem) and describe its necessary and desirable features. Finally, the results of our analysis show that our proposed ecosystem surpasses the existing ecosystems for U-VR applications and contents.

1 Introduction

An ecosystem is a system defined by the interactions of a community of organisms with each other and with their environment; in a natural ecosystem, the natural environment is in which interdependent organisms, such as plants and animals, and physical factors of the environment, such as rocks and soil. [1] The interactions between the organisms in computing resources (e.g., devices, systems) including users can define non-natural ecosystems such as digital, mobile, or content ecosystems.

Many researchers describe digital ecosystems [2-5], defined as ecosystems formed by computing resources (e.g. devices, services, data), in order to acquire new perspectives on information systems. Some researchers focus on mobile ecosystems [6], and more rarely on content ecosystems. Mobile ecosystems are digital ecosystems in which the mobility of users drives interactions and evolutions. Typically, a mobile device informs a service about its position, orientation or movement, and the service processes this information to suggest or provide appropriate contents. For example, a smart phone may use an embedded GPS to tell its position to a navigation service, which calculates and displays a path to the user's destination, indicating restaurants on the way. Besides, content ecosystems are digital ecosystems that preserve the order of

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 377–386, 2011.

contents that are created, maintained, modified, evolved, or destroyed by means of users' participation. Typically, the devices provide stable standard hooks (e.g. API, codec), and are considered more or less equivalent from a systems perspective; rather than organisms, they become parts of the environment (which does not preclude improvements).

Mobile ecosystems and content ecosystems can both be extended but are basically disjoint; their features can promote services in ubiquitous virtual reality but are difficult to combine in their current form. Consequently, it is important to define and realize a new type of ecosystem for u-Contents (u-Content ecosystem), based on mobile and content ecosystems, ideally with minimal fundamental changes to effectively and efficiently reuse existing works. Here, we use u-Contents according to the following definition: “contents appropriate for an UbiComp-enabled smart environment, making virtual reality pervasive into our daily lives” [7]. The u-Content highlights the importance of users, who are continuously immersed in a mix of real and virtual spaces (aka U-VR [8] space), whose contexts and activities influence the relevance of contents and services, and who may produce as much as consume contents.

For the u-Content ecosystem, we should precisely match contents to users' characteristics (e.g. abilities, preferences, viewpoints), to the physical world (e.g. location, luminance, noise), events, social situations, and to goals of producers and consumers. We also need to motivate and support a sustainable participation of users as mobile producers and consumers, which implies simple contents creation, flexible distribution, easy retrieval, and seamless interaction. Finally, the permanent self-growth of contents is highly desirable.

Through this paper, we consolidate information from diverse sources, providing a consistent bibliography from fields such as information systems, human-computer interaction, and the semantic web. Our survey revealed 8 key concepts: digital food (resource needed by contents), digital organ (component of contents), digital species (type of contents on various platforms and devices), digital community (group of contents), digital environment (e.g. mobile device, intranet, Internet), digital ecosystem type (e.g., mobile, contents, digital health ecosystems), eco-relationship (relationship between two or more digital species), and eco-law (e.g. format, application specification and restriction). Based on our thorough analysis using the key concepts, our taxonomy helps visualize all related aspects and find out essential technical components for u-Content ecosystems.

Moreover, we demonstrate that traditional digital ecosystems cannot appropriately process u-Contents due to low considerations for human factors (e.g. restrictions for kids, adaptations for novices, rewards to motivate end-user participation), energy management (e.g. contents download consumption), and users' context (e.g. profile, user-created contents, location, time, activity, belonging to a community). Our analysis also highlights weaknesses in the temporal management (up-to-date/real-time, old/archived, future/predictions), in the contents management (equal versus unequal treatment), and in the provision of alternatives to find, combine (e.g. subtitles, translations), transform (e.g. format, version), and use contents.

In this paper, we propose the foundation of a new digital ecosystem for u-Content. First, we survey the digital ecosystems, explore their differences, identify unmet challenges, and consider their appropriateness for emerging services tightly linking real and virtual (i.e. digital) spaces. Second, we define a new type of content ecosystem (u-Content ecosystem) and describe its necessary and desirable features.

The proposed u-Content ecosystem enables new services while facilitating the realization–and enhancing the quality–of existing services. Elements of interest concerning our u-Content ecosystem notably include the provision of adaptation mechanisms to enable personalization and situational focus, and of reward mechanisms for content contributors to encourage participation. Finally, the results of our analysis show that our proposed ecosystem surpasses the existing ecosystems for U-VR applications and contents.

2 Key Concepts and Analysis

In this section, we define, clarify and exemplify the 8 key concepts that our survey revealed: digital food, digital organ, digital species, digital community, digital environment, digital ecosystem, eco-relationship, and eco-law. We then analyze related literatures by applying the key concepts and sometimes we interpret the literatures based on our view.

2.1 The 8 Key Concepts of the Digital Ecosystem

Digital foods are resources individuals must consume to survive and prosper [3], including data, memory space, and processing cycles. Here, “consume” means various things: data may remain available without change, memory may be freed later, and processing cycles may be replenished.

Digital organs are permanent software and hardware parts of individuals that support specific features of digital species [2]. Mobile ecosystems typically feature antennas and algorithms for power management, localization and face recognition; content ecosystems notably feature affective meta-data [9].

Digital species mimic biological species, living organisms that are autonomous, viable, and self-organizing, based on purposes, hardware and software [2, 3]. Alternatively, some consider the abstractions of the devices as species, instead of the devices themselves [3]. In any case, digital species may be neutral to, benefit or harm their ecosystem. In generic ecosystems, digital species notably include computers, displays, cameras, smart cards, and card readers; in specific ecosystems such as health ecosystems, they may be medical devices or measuring tools [2]. In content ecosystems, digital species notably include 3D models and 3D contents.

Digital communities are groups formed by digital species, mimicking communities of intelligent or stupid animals and plants [2].

Digital environments are environments in which digital species evolve, and in which individuals complete their life-cycle. Environments apply pressure but also offer cradles, dwellings, graves, and resources (e.g. information and transaction flows [2]). Digital environments may be spatial [3], with functions to indicate local resources and neighbors, or not; for example, environments for virtual reality are typically spatial whereas the World Wide Web is basically non-spatial. Ecological niches associated to idiosyncratic eco-laws may form the environment [3], with a dynamic hierarchy of niches containing one or more niches (enclosing niches) and fully contained niches (enclosed niches). In generic ecosystems, nested ecological niches may be printers in smart rooms in smart buildings in a smart city. In health ecosystems, information includes personal health records, and transactions include verifications and money transfers when purchasing drugs [2].

A digital ecosystem is a "dynamic and synergetic complex of digital communities consisting of interconnected, interrelated, and interdependent digital species situated in a digital environment that interact as a functional unit and are linked together through actions, information, and transaction flows" [2]. Examples include content ecosystems, mobile ecosystems, pervasive ecosystems [3], health ecosystems [2], and cyber-physical ecosystems such as the Energy Web [10].

Eco-relationships are relationships between digital species such as commensalism, mutualism, parasitism, and predation. Although central in Nature, predation is rarely presented in digital ecosystems; for examples of digital predation, see [3]. Hadzic and Chang [2] defines digital symbiosis as similar to biological symbiosis, with 3 types: mutualism (both species benefit), commensalism (one species benefit with no cost to the other) and parasitism (one species benefits at cost to the other). For example, digital symbiosis can appear between decision-support and reservation systems [2].

Eco-laws are fixed permanent laws regulating local interactions and changes in a digital ecosystem [3, 10]. They may determine how and when individuals can reproduce, find/affect/kill individuals, or produce/find/alter/consume resources [3]. Villalba et al. [3] distinguishes internal from external laws, managing what happens respectively within a niche and between a niche and the rest of the space. Elsewhere, the laws may be phylogenetic (i.e. linked to the evolution of species), ontogenetic (i.e. linked to individual development, with little environmental influence) or epigenetic (i.e. linked to learning from environmental interactions) [10].

Finally, humans and communities thereof strangely fit digital ecosystems: people can be seen on a par with digital species or environmental resources... or may be ignored altogether. When considered, their main generic categories are producer [3] (e.g. designer, advertiser [11], service developer [3]), consumer [3, 10] (e.g. audience [11]), prosumer [10], and maintainer/owner (e.g. display owner [11]).

2.2 Analysis of the Survey

One of examples for digital foods is museum contents which include objects like images and their metadata for tagging and annotating [12]. That is, different groups of stakeholders (museum curators, owners, professional researchers and the general public) add new objects and attributes, edit, and browse the collection individually. Another example of digital foods is contents/articles which are posted on a web-page (i.e. Slashdot.org) for idea sharing. Slashdot is a virtual community site where community members write articles. These articles usually comment on current affairs, such as a report or news item written elsewhere, which the author then critically summarizes and links to [13]. Liu et al. [14] proposed designed and implemented a robust and adaptive super-peer based algorithm to construct and manage the digital ecosystem where each digital device (node) sends service requests to share resources. We can consider these kinds of service requests and services themselves as digital foods. Hadzic and Chang [2] defined digital organs as software and hardware that will support the digital species features.

Boley and Chang [15] defined agents as entities that join an environment or a community based on their own interests, and species as types of agents, respectively. Hadzic and Chang [2] described digital organs are assembled into a digital species body. They mentioned the majority of digital species consist of 'hardware' together with its associated 'software'. Briscoe [16] mentioned that the agents of the digital ecosystem are functionally parallel to the organisms of biological ecosystems, including the behavior of migration and the ability to be evolved. In [12], we can think of various management interfaces, such as the tagging and browsing software and associated hardware, as digital species. That is, they provide different user interfaces for adding/editing new objects and attributes, and for browsing/searching the collection. It seems that Wagner et al. exemplified digital species as a unique online comment moderation process engine that is "slash" engine. Anyone can comment, even anonymously, but comments are subsequently rated for quality for balancing egalitarian forces with quality assurance [13]. In [14], a large collection of digital devices and super-peer based algorithm to construct and manage the digital ecosystem overlay could be digital species. Digital devices communicate through message exchanges, and every digital device is a node in the digital ecosystem overlay.

Even though Boley and Chang [15] mentioned the concept of digital communities, those are very similar to the description of a digital environment. Hadzic and Chang [2] also considered digital communities as some groups in a digital environment. Briscoe [16] simply defined a digital community as strongly connected cluster or group based on [17] where Begon et al. defined a community as a naturally occurring group of populations from different species that live together, and interact as a self-contained unit in the same habitat which is a distinct part of the environment. In the virtual museum application [12], digital community/environment can be considered development platform relying on web technologies (including web 2.0 and web services) and semantic web technologies.

Boley and Chang [15] defined an (agent) environment as an environment which contains human individuals, information services as well as network interaction and knowledge sharing tools along with resources that help maintain synergy among human beings or organizations. The definition of a digital environment by Hadzic and Chang [2] is an environment in which digital species jointly live, function, and relate. In addition, they mentioned that the digital environment is very analogous to the biological environment, and the digital environment was considered as an important constituent for digital ecosystem. Briscoe [16] considered an environment as population's habitat where different species live together and interact.

Digital ecosystems are self-organizing systems which can form different architectural models through swarm intelligence, where local interactions between agents determine the global behavior [15]. Hadzic and Chang [2] defined a digital ecosystem as the dynamic and synergetic complex of digital communities consisting of interconnected, interrelated, and interdependent digital species situated in a digital environment that interact as a functional unit and are linked together through actions, information, and transaction flows. Briscoe [16] considered digital ecosystems software systems that exploit the properties of biological ecosystems. Eklund et al. [12] designed a digital ecosystem of the virtual museum of the pacific, which is a web-based knowledge management system (or information space) for ethnographic objects. Wagner et al. [13] defined a digital ecosystem as a kind of community where different parties with different levels of interest and resources participate and provide resources in a give-and-take that creates benefits for all. A digital ecosystem is an emerging paradigm for economic and technological innovation to support the cooperation, knowledge sharing and development of open and adaptive technologies [18].

It seems that Boley and Chang [15] emphasized the benefit and profit of species through the relationship within communities. Precisely, collaboration between entire digital ecosystems should be able to provide benefits to all ecosystems involved. In addition, participation of individuals in multiple digital ecosystems helps them to reach their goals. Eklund et al. described relationships across ecosystems as any inferences that are formed as a result of user communities tagging and annotating objects (or writing wiki entries) in their own group [12]. For example, historians, anthropologists and curators are each members of their own 'group'. If anthropologists are a group within their own 'ecosystem' and the historians belong to another, a cross-ecosystem relationship is formed from their respective collaborative inputs into the virtual museum. Wagner et al. [13] stressed on a dynamic give-and-take between community members to achieve benefits for all through an effective information and coordination mechanism, those also require little learning and little effort to maintain, i.e., a well designed system enabling the professional virtual community. Sometimes, species cooperate with each other in order to share resources [14]. They dispatch the service requests from users to client nodes. One peer group is able to provide a type of service that differs from the other peer groups.

An example of eco-law could be the acceptance and rejection process by moderating editors [13]. After articles are posted by community members on the virtual community site, there is a moderation process for each article.

3 Ecosystem Realizations

To realize ecosystems, we may integrate many complementary concepts and technologies.

3.1 Necessary and Desirable Features

In the case of content ecosystems, the Semantic Web techniques (based on formal syntaxes, ontologies, inference rules) can greatly support human-machine and machine-machine interoperability. For example, we can consider discovery process based on semantic matching within a niche (or nearby one) [3]. Besides, Villalba et al. [3] identifies a single key component for passive species: a public semantic description of its features; active species add: a set of needs e.g. nature of its foods, a happiness status driving activities, and an internal logic to maximize its happiness. Here, happiness is typically goal/need-oriented and time-dependent, and may depend on the happiness of nearby individuals. These components may notably enable evolvability and self-healing but may hamper privacy awareness and self-protection. Finally, individuals/species should ultimately be endowed with manipulation to reconfigure themselves and the (real) world/environment [19]. This ability would most notably increase robustness and self-protection while uniquely helping deal with limited access.

3.2 Application

This section exposes the scope and interest of attempts to establish ecosystems so far and in particular content ecosystems. The concept of ecosystem has been applied to the domains of advertisement, energy and health. Besides, Hadzic and Chang [2] suggested digital business ecosystems, digital government ecosystems, digital law ecosystems, and digital industry ecosystems but without clarifying their potential components.

In the domain of advertisement, Villalba et al. [3] discussed an advertising ecosystem. For example, in MyAds [11], information and advertising agents could move among screens, looking for people with specific user profiles, and then try to control a selected screen with their own contents [3]. In the "display niches", information feeds user agents, who feed advertiser agents then display-bound monitoring agents selectively spread information and guide advertiser agents, forming a feedback loop; eco-laws define how user agents make user profiles available in the displays [3]. In cloud-based ad printing service [20], web content is automatically reproduced with automatic contextual ad and coupon insertion when the web content is printed.

In an Energy Web system [10], autonomous agents control the production and usage of distributed renewable energy sources to optimize energy use and allow rapid innovation. It is envisioned based on 5 layers: distribution, control, prosumer, community, and incentive (from more technical to more social). A digital health ecosystem has been proposed, focusing on electronic health records, notably involving hospitals, health services, general practitioners, pharmacies, health systems, and health information resources [2]. According to [19], ecosystems may increase the efficiency of, and innovation in, mHealth (e.g. disease prevention and management) by providing flexibility.

In the domain of P2P applications, the BitTorrent ecosystem is one of the successful open internet applications in the present. Zhang et al. [21] crawled popular torrent-discovery sites than studied in-depth the Ecosystem’s tracker, peer, user behavior, and content landscapes. In the domain of digital museums, information and knowledge of museum artifacts have been digitally collected to construct the ecosystem of the virtual museum [12]. Additionally it encourages sustainable user participant by containing a wiki component, where the ecosystem enables the creation of user-defined perspectives with semantic theme.

All these ecosystems could be merged into a universal digital ecosystem [2], in the same sense that our natural ecosystems can be merged into an Earth ecosystem. Conversely, many digital ecosystems could be split into lower-level ecosystems; for example, a digital ecosystem may be split into a digital dental ecosystem, a digital physiotherapy ecosystem, a digital surgery ecosystem, and so on [2].

4 Discussion

Considering u-Content and its characteristics, the existing ecosystems have strong limitations linked to contents manipulation, privacy, standards, user participation, and management.

Contents are still difficult to manipulate, notably because of structures (e.g. file formats) are not organized to allow innovations and because semantic information is missing. However, solutions can be integrated in ecosystems. For example, to facilitate pervasive video editing, Paluska et al. [22] proposed the concept of *chunk*: a "typed, ordered, fixed-size array of fixed-size slots" with slots holding scalar data and linking to another chunk. *ChunkStream* demonstrated adaptations to local constraints [22]: bandwidth using a search tree, and screen sizes using several fidelity levels. These chunks can be generalized to other types of applications and contents. Regarding semantic information, many types of contents lack semantics, techniques to provide semantics, and tools to enable interoperability; for instance 3D media [23], sounds and smells. For u-Content manipulation within the ecosystem, representations of u-Reality, u-Intelligence, and u-Mobility need to be formally defined. Metadata that describe u-Content can be further utilized in means of context-aware manipulation for filtering and recommendation. With this support, u-Content can be updated and modified by multiple users and enriched collectively by users as well as between interactions of u-Contents.

u-Content may be particularly constrained legally because, by their nature, they heighten privacy threats. For example, Europe's rights commissioner discussed in 2010 a legal "right to be forgotten" online (correction, withdrawal, deletion of contents) that may be included in a soon-to-come update of the EU data protection rules [24]. Existing ecosystems do not particularly facilitate contents removal. Ecosystem features that regulate what to share, how to share, who gets the right/ownership and granting accessibility of who can view, edit and publish need to be included in u-Content ecosystem.

Ecosystems typically lack standardized formats and (high capacity, quality) infrastructures to share contents, which is particularly true for health services [2]. This is also true for u-Content but the difference is that u-Contents are just emerging: now is the best moment to establish standards, now that we can see what u-Content should be, and before specific solutions become de facto standards. Such activities are currently taking place in u-Content [7] and mixed reality content representation [25].

User participation is an important feature missing in usual digital ecosystems. A user may stop providing contents or feedback if her costs and expectations exceed her benefits. For participatory sensing, costs include bandwidth, CPU and energy consumption as well as privacy [8]. For user participation, Lee et al. [8] proposed a *Reverse Auction based Dynamic Price incentive mechanism with Virtual Participation Credit*, maintaining a sufficient number of users providing environmental data while lowering costs to purchasers. Provision of incentives to users as well as providing gaming and engaging experiences should also be included in u-Content ecosystem.

Finally, we lack efficient techniques to manage distributed online evolution [10], suitable ways to measure fitness in a distributed way with noise [10], suitable metrics to assess robustness and resilience in autonomic complex systems [10], guidance for diagnosing e.g. simulations to test ecosystems and security assessments. Villalba et al. [3] even points out that we ignore what the threats are in a digital ecosystem.

Considering the properties we wish to imbue, and the limitations noted so far, it appears that inventing one or several appropriate frameworks will require direct hands-on experience.

5 Conclusion

In this paper, we analyzed digital ecosystems to demonstrate the need for a new digital ecosystem, oriented towards contents for ubiquitous virtual reality. We surveyed the digital ecosystems and defined a new type of content ecosystem (u-Content ecosystem). Through this paper, we consolidated information from diverse sources, providing a consistent bibliography from fields such as information systems, human-computer interaction, and the semantic web. Based on the results of our analysis, we showed that our proposed ecosystem surpasses the existing ecosystems for U-VR applications and contents. We expect our proposal to lead to self-emergence at both the micro (better contents and more interesting links between contents) and macro (prosumers) levels.

References

1. Wikipedia, `http://en.wikipedia.org/wiki/Ecosystem`
2. Hadzic, M., Chang, E.: Application of Digital Ecosystem Design Methodology Within the Health Domain. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 40(4), 779–788 (2010)
3. Villalba, C., Mamei, M., Zambonelli, F.: A self-organizing architecture for pervasive ecosystems. In: Weyns, D., Malek, S., de Lemos, R., Andersson, J. (eds.) SOAR 2009. LNCS, vol. 6090, pp. 275–300. Springer, Heidelberg (2010)
4. Dong, H., Hussain, F.K., Chang, E.: Semantic Service Retrieval and QoS Measurement in the Digital Ecosystem Environment. In: International Conference on Complex, Intelligent and Software Intensive Systems (CISIS), pp. 153–160 (2010)
5. Pfeifer, T., Malone, P.: Digital Ecosystem for Cross-domain Context Sensing and Tradinz. In: 32nd IEEE Conference on Local Computer Networks (LCN 2007), pp. 228–230 (2007)
6. Xia, R., Rost, M., Holmquist, L.E.: Business Models in the Mobile Ecosystem. In: 2010 Nineth International Conference on Mobile Business / 2010 Ninth Global Mobility Roundtable, pp. 1–8 (2010)
7. Kim, K., Han, J., Kang, C., Woo, W.: Visualization and Management of u-Contents for Ubiquitous VR. In: The 14th International Conference on Human-Computer Interaction (2011) (to be published)
8. Lee, Y., Oh, S., Shin, C., Woo, W.: Ubiquitous Virtual Reality and Its Key Dimension. In: International Workshop on Ubiquitous Virtual Reality, pp. 5–8 (2009)
9. Tkalčič, M., Burnik, U., Košir, A.: Using Affective Parameters in a Content-based Recommender System for Images. User Modeling and User-Adapted Interaction 20(4), 279–311 (2010)
10. Carreras, I., Miorandi, D., Saint-Paul, R., Chlamtac, I.: Bottom-Up Design Patterns and the Energy Web. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 40(4), 815–824 (2010)

11. Di Ferdinando, A., Rosi, A., Lent, R., Manzalini, A., Zambonelli, F.: MyAds: A System for Adaptive Pervasive Advertisements. Pervasive and Mobile Computing 5(5), 385–401 (2009)
12. Eklund, P., Wray, T., Goodall, P., Bunt, B., Lawson, A., Christidis, L., Daniels, V., Van Olffen, M.: Designing the Digital Ecosystem of the Virtual Museum of the Pacific. In: 3rd IEEE International Conference on Digital Ecosystems and Technologies, pp. 805–811 (2009)
13. Wagner, C., Liu, L., Schneider, C., Prasarnphanich, P., Chen, Huaping.: Creating a Successful Professional Virtual Community: A Sustainable Digital Ecosystem for Idea Sharing. In: Digital Ecosystems and Technologies (DEST 2009), pp. 163–167 (2009)
14. Liu, M., Koskela, T., Ylianttila, M.: Constructing a Robust Overlay for Digital Ecosystems. In: Digital Ecosystems and Technologies (DEST 2009), pp. 430–435 (2009)
15. Boley, H., Chang, E.: Digital Ecosystems: Principles and Semantics. In: Digital Ecosystems and Technologies (DEST 2007), pp. 398–403 (2007)
16. Briscoe, G.: Creating a Digital Ecosystem: Service-Oriented Architectures with Distributed Evolutionary Computing. In: Java One Conf., p. 0759 (2006)
17. Begon, M., Harper, J., Townsend, C.: Ecology: Individuals, Populations and Communities. Blackwell Publishing, Malden (1996)
18. Nachira, F., Dini, P., Nicolai, A., Le Louarn, M., Rivera Lèon, L.: Digital Business Ecosystems. The European Commission (2007)
19. Estrin, D., Sim, I.: Open mHealth Architecture: An Engine for Health Care Innovation. Science 330(6005), 759–760 (2010)
20. Yang, S., Jin, J., Parag, J., Liu, S.: Contextual advertising for web article printing. In: The 10th ACM Symposium on Document Engineering (DocEng 2010), pp. 195–198 (2010)
21. Zhang, C., Dunghel, P., Wu, D., Ross, K.W.: Unraveling the BitTorrent Ecosystem. IEEE Transactions on Parallel and Distributed Systems (2010); Online published (10.1109/TPDS.2010.123)
22. Paluska, J., Pham, H., Ward, S.: ChunkStream: Interactive Streaming of Structured Data. Pervasive and Mobile Computing 6(6), 607–622 (2010)
23. Spagnuolo, M., Falcidieno, B.: 3D Media and the Semantic Web. IEEE Intelligent Systems 24(2), 90–96 (2009)
24. Waterfield, B.: Right to be forgotten. The Telegraph (2011)
25. Kim, H., Yoon, H., Choi, A., Baek, W., Lee, I., Kim, D., Woo, W.: Data Markup Representation for Mixed Reality Contents. In: International AR Standards Meeting 2011 (2011)

Semantic Web-Techniques and Software Agents for the Automatic Integration of Virtual Prototypes

Rafael Radkowski[1] and Florian Weidemann[2]

[1] Heinz Nixdorf Institute, Fürstenallee 11, 33102 Paderborn, Germany
[2] University of Paderborn, Warburger Straße 100, 33100 Paderborn, Germany
rafael.radkowski@hni.uni-paderborn.de, fweidema@uni-paderborn.de

Abstract. A virtual prototype is a computer internal representation of a real prototype. It is composited by a set of different aspect models. A common technique to analyze a virtual prototype is the usage of virtual reality applications. However, this requires a composition of different aspect models and their integration into a virtual environment. In this paper, an agent-based technique is presented, which facilitates this automatic integration of different aspect models. The Resource Description Framework is utilized to annotate the aspect models. A software agent compares the annotations of different models. By this it identifies similar models. A software prototype has been created that shows the usefulness of the approach.

1 Introduction

A virtual prototype (VP) is the computer internal representation of a real prototype or product [1]. It is based on the digital mock-up (DMU). The DMU represents the shape and the structure of the product. Two types of models are the origin of the DMU: 3D-CAD models and the logical structure of the product. The VP extends the DMU by further aspects. Aspects like the kinematic, dynamic, tension or information processing. A computer-internal model represents each of these aspects [2].

For the analysis of VPs the technology virtual reality (VR) and VR-applications are utilized commonly. VR displaces the user into a three-dimensional, computer-generated virtual environment (VE) [3]. To analyze a virtual prototype in the VE, the VP is integrated as a 3D model. Aspects like the kinematic behavior, tensions, etc. are calculated using simulation software. The results of these calculations also have to be integrated into the VE and to be combined with a certain VP. Therefore, it is necessary to combine different models, which represent the different aspects of a VP. The relations between the aspect models have to be identified, the data has to be specified that describe these relations, and finally everything have to be implemented. Today, different engineers do this integration manually, supported by several graphical interfaces that visualize the relations.

However, the manual integration is error-prone and time-consuming. The engineers have to know very well the meaning of different models. This is difficult when they do not know these models or use black-box models. Furthermore, the manual integration is labor consuming. Meanwhile, the engineers cannot focus on their intrinsic engineering tasks.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 387–396, 2011.

To face this problem, the Collaborative Research Center (CRC) 614 “Self-optimizing Concepts and Structures for Mechanical Engineering” follows a novel approach: Aspect models of VPs are annotated by techniques commonly used in the field of the semantic web. Software agents use these annotations in order to composite VPs autonomously and finally, they integrate them into a VE.

This paper is structured as following: The next section introduces some related work. The concept is explained in section three. Then an application example is presented. The paper closes with a summary and an outlook.

2 Related Work

The related work is separated into two sections. First the related work in the field of software agents supported engineering is reviewed. The second part introduces the utilization of the Resource Description Framework (RDF) to model technical systems.

In engineering, software agents are utilized in many different application fields. Mainly agents are used to support the design process by making decisions, which are based on a large amount of data. Mendez et al. describe an agent-based software architecture for agents in virtual environments [4]. They introduce the concept of expert agents. Expert agents are software agents with an expert knowledge in a specific technical domain. Based on this knowledge, the expert agent is capable to find a solution to solve a specific problem. This paper introduces a similar idea. However, their desired tasks are training tasks. Galea et al. present a framework for an intelligent design tool that assists a designer while working on micro-scale components [5]. They do not label their framework as software agent, but they use a similar artificial intelligence technique to model the knowledge and the reasoning system. Multi-agent systems have also been used to support engineers in time-critical tasks [6]. An agent aggregates relevant information from other agents that represent different members of an engineering team. Thus, an engineer gets the right information at the right time. Baolu et al. propose the so-called Multi-Agent Cooperative Model (MACM) [7]. It is a product design system that allows an easy access to similar data of different products. The system facilitates the product design and manages product data. With its aid the product design cycle will be shortened. Geiger et al. introduce the agent modeling language SAM (Solid Agents in Motion), a language to describe 3D models in virtual environments and their behavior. This work is similar to our approach. However, the agents generate rule-based animations to explain the kinematic behavior [8].

Some researchers have already used RDF and the related reasoning mechanism for the engineering of technical systems. For instance, Bludau and Welp have developed a framework, which supports engineers during the development of mechatronic systems [9]. Their framework searches for active principles and solution elements, which meet a given specification. Restrepo uses Semantic Web (SW) techniques to search for design solutions for a given problem [10]. He has developed a database, which contains different design solutions; every solution is annotated by RDF. A reasoning mechanism searches for solutions, which meets the given design problem. Ding et al. use XML-based annotations to annotate CAD models with design constrains, goals, relationships, and bounds [11]. They mainly annotate geometric, topological and kinematic properties of a given design. Their approach can be utilized to find an optimal design solution during the product development process. The authors use XML as notation basis, but their notation is similar to a RDF notation. Ding et al. developed an XML-based product representation that allows an annotation of geometrical properties, too [12]. For further information, Li et al. present a classification of different annotation approaches [13].

In summary, the work supports the importance of software agents and annotation techniques in engineering design.

3 Concept for the Automatic Integration of Virtual Prototypes

Fig 1 shows an overview of the concept. On the bottom of the figure, the grey circle indicates the virtual environment. The boxes inside the virtual environment symbolize the VPs. A software agent represents each VP.

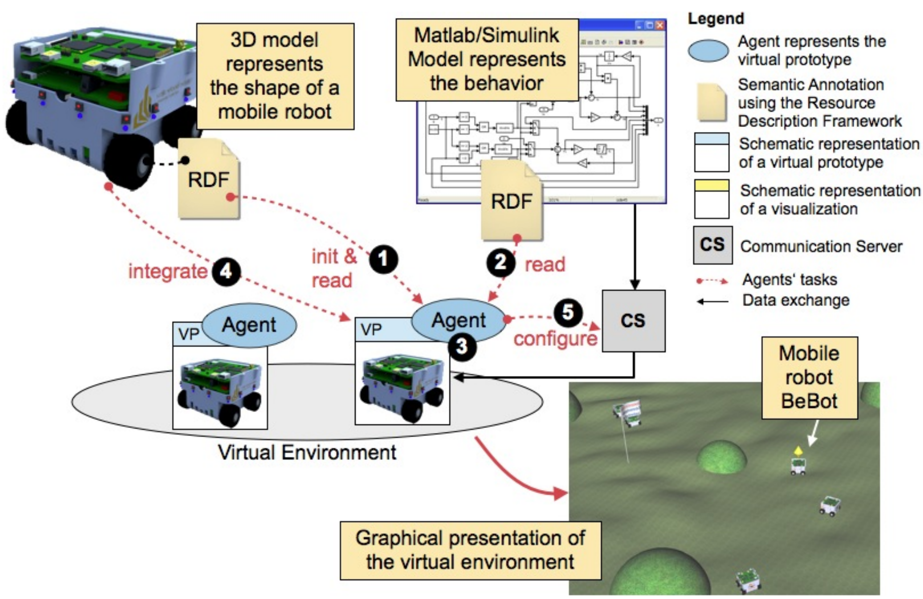


Fig. 1. Overview of the concept

In this example, a VP includes two models: a 3D model and a behavior model. Both models are shown on top of the figure. Left, the 3D model, the behavior model on the right. In this case it is a screenshot of Matlab/Simulink. The application contains and processes a model that simulates the behavior. Both models are annotated. Therefore, an RDF-notation is utilized; the annotations describe the purpose of the models. Normally, more than two aspect models (3D model, behavior) and one VP are used. The small example should facilitate the understanding of the concept.

Objective of the software agent is to combine both aspect models (3D model, behavior) to one virtual prototype and to integrate them into the VE. In order to realize this a VP-template exits inside the virtual environment. The task of the agent is to integrate the aspect models into that template. It proceeds as following: The first step is an initialization by a user (1). Normally, the user specifies one model (3D model or behavior model) as origin. The objective of the agent is to identify the other models and to integrate them into the template of the VP. Therefore, it searches for every available model. A service directory of the agent platform references them. The annotation of every available model is read (2). The agent compares the RDF model of the 3D model with the RDF model of the behavior model (3). A set of production rules is used for this task. If two models pass the production rules, the agent assumes them as similar.

Then the 3D model is integrated into the VE by loading the model and including it into an internal data model (4). The behavior model cannot be included by loading. Normally, its processing would exceed the resources of every computer very fast. The behavior model is executed on a different computer system. However, only the results of the simulation are necessary for an analysis of the VP. Therefore, the results need to be transmitted to the VP. A communication server (CS) is used for that purpose. It manages the communication between the simulation software and the VP/VE. The task of the agent is to establish the communication between the behavior model and the 3D model. Therefore, the agent has to configure this communication server (5).

3.1 Semantic Annotations with the Resource Description Framework

The Resource Description Framework (RDF) is a description language, which is used to annotate the content of a web page in the context of the Semantic Web [14]; it is a syntax for meta data of a web page. The underlying model is based on a directed graph. The nodes of the graph are denoted as resources, the edges are denoted as properties. The idea of RDF is to describe complex facts by a network of simple RDF statements. A RDF statement consists of a subject, a predicate and an object.

In our concept the semantic annotation with RDF describes the purpose of each model. The challenge during the generation of the annotation has been to figure out the elements of a certain model, which need to be annotated in order to describe the purpose of a model and finally to facilitate the automatic integration of the model. We have specified models for following aspects of a VP: shape (3D model), behavior, functions, and activates. In the following the annotation of a 3D-model is explained by an example. **Fig 2** shows a 3D model of a shock absorber and its annotated items. The symbols in the figure show a graphical representation of an RDF-based semantic annotation.

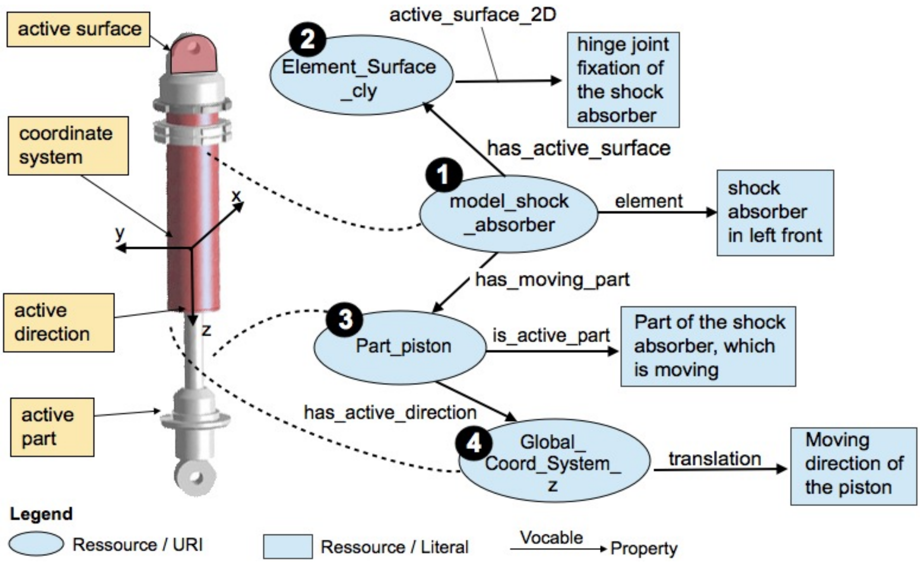


Fig. 2. Example of the semantic annotation of a 3D model

Four items of a 3D model has to be annotated: the entire part, the active surfaces, the sub-parts and the active directions:

- Entire part (1): The resource is linked to the variable, which represents the model, normally a file. In this example, the name of the model is *model_shock_absorber*. The variable is annotated by the predicate *element*. To describe the *element*, a literal is used. In this example it describes the part as “shock absorber in front left”.
- Active surface (2): The active surfaces of a component are the surfaces, which fulfill the functions of this component [15]. The resource is linked to the variable in the data structure of the 3D model, which represents the active surface. In the example, the name of the variable is *element_surface_cly*. As predicate the word *active_surface_2D* is used. This predicate specifies the item as active surface.
- Active part (3): This type of part is moving to cause an effect of the entire model. The resource refers to the variable, which describes the main part in the data structure of a 3D model; here it is the entire pisiton. The word *has_moving_part* is used as RDF predicate to annotate the sub-part, which describes the part in the structure of the entire 3D model; the variables name is *part_piston*. Furthermore, to describe this active part, a predicate *is_active_part* is used. It facilitates the annotation with a literal. In this case the literal contains: “Part of the shock absorber, which is moving.”
- Active direction (4): The fourth annotated type of item is the active direction. According to Pahl/Beitz [15] the active direction describes the direction, into which a function of a component effects. In figure 2, the piston of the shock absorber is the active part, which active direction should be described. The variable *global_coord_system_z* describes the coordinate system; it describes the direction of moving. It is attached to the resource *part_piston* by the predicate

has_active_direction. In addition, the resource *global_coord_system_z* needs to be annotated with a human understandable literal. In the example, the literal says "*Moving direction of the piston*". It is attached to the resource by the predicate *translation*.

The entire description shows one example only. The example should give an impression, how the semantic annotation with RDF is working and which elements of a 3D model are necessary in order to describe the purpose and functionality of a 3D model in a natural way (literals). Altogether 36 RDF keywords were defined to describe the active surfaces and directions as well as the parts / sub-parts of an assembly. Further details of the model are explained in [16].

3.2 Software Agent Reasoning

The software agent has two major tasks: First, it has to find similar aspect models and second it has to establish the communication and the exchange of data between different software tools. In the following the reasoning mechanism is described. **Fig 3** shows the principle by an example.

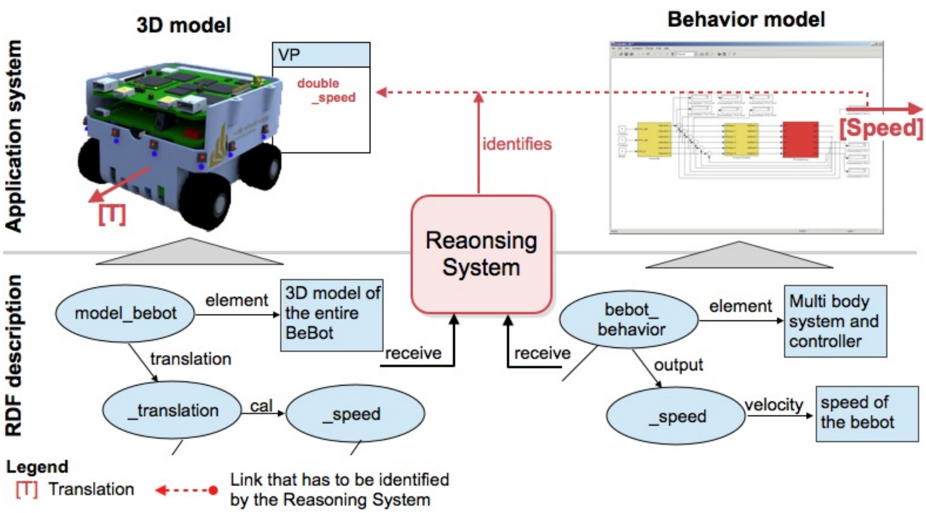


Fig. 3. Schematic overview of the reasoning by an example

On the left side of the figure a 3D model of a mobile robot is visualized, on the right side the behavior model of the robot is indicated. Both models are annotated using an RDF notation; the figure shows a part of the RDF description only. The mobile robot is the so-called BeBot, which is under development at the Heinz Nixdorf Institute. The BeBot is an autonomous driving robot; it can fulfill different tasks in a team. To get a VP of the BeBot the variables of the behavior model need to be linked with the related variables of the 3D model. In this example the variable *speed* is shown in the figure.

The reasoning mechanism identifies variables that are related to each other. In general the software agent compares the RDF models, two at the same time, and converts the results of the comparison into a numerical value. This numerical value expresses the similarity of two models respectively their variables. The comparison is based on production rules. Each production rule has the form:

$$\text{IF (Condition } C_1 \text{ \& Condition } B_1 \text{ \&} \ldots \text{ \& Condition } C_n \text{ \& Condition } B_m) \text{ THEN } A_1; \ldots; A_o \quad (1)$$

Conditions of type *C* are predicates of the 3D model, conditions of type *B* are predicates of the behavior model. By these production rules a set of corresponding predicates is identified. As corresponding predicates each pair of predicates is defined, which describes the same meaning of an item. For instance, condition *C* states *translation & cal* (calculate) and condition *B* states *output & velocity* are defined as corresponding predicates; they result in an output $A = 1$. Otherwise they result in an output $A = 0$. The result of this calculation is weighted by a weight value *g*:

$$A_o = A\,g + E \quad (2)$$

The value *g* indicates the importance of a production rule. The term *E* is an offset. It is calculated by a comparison of the literals of each pair of corresponding predicates. This is done by a statistical phrase analysis (for further details [16]). The results of every production rule are described as a vector:

$$A_{similar} = \{A_1, A_2, \ldots, A_o\} \quad (3)$$

This vector is a rating scale for the quality of the similarity in a certain task.

After the vector is determined the agent ranges all results $A_{similar,i}$, where the index *i* refers to a certain production rule of two compared models. A statistical method is used for this comparison, the so-called squared ranking. This method calculates a likelihood value *p(i)* for each corresponding pair of predicates:

$$p(j) = \frac{1}{size} \cdot \left(E_{\max} - \left(E_{\max} - E_{\min}\right) \cdot \frac{(R_{similar,j} - 1)^2}{size - 1} \right) \quad (4)$$

With two rating values E_{max} and E_{min}. These values express the estimated amount of minimal and maximal corresponding predicates, respectively the number of possible relations. The equation assigns a numerical value to each production rule and expresses the fulfilled rules by a numerical value. A high value indicates the similarity of the compared variables.

The agent links all data, whose value *p(j)* exceeds a threshold $p_{threshold}$.:

$$p(i) > p_{threshold}$$

At this time the threshold is determined empirically. After this decision, the agent establishes the communication between the behavior model and the 3D model. Further information about the communication infrastructure and the behave of the agent inside the virtual environment has been presented in [17].

4 Application Example

To test the entire concept a software prototype has be developed and the described concept has been realized and integrated into this application. The software prototype consists of four components: The first component, a virtual environment is based on OpenSceneGraph (www.openscengraph.org), an open source scenegraph library for the development of 3D graphic applications. The second component is a simulation for mobile robots. This simulation is carried out with Open Steer (http://opensteer.sourceforge.net/), an open source software library. The third component is JADE (Java Agent DEvelopment Framework). JADE is a software framework that facilitates the implementation of multi-agent systems through a middleware that complies with the FIPA (Foundation for Intelligent Physical Agents) specifications, a standard specification for software agents. The fourth component is a communication server. It realizes the exchange of data between the three components, mentioned before. The technical aspects of the server are described in [17].

A test application has been developed to analyze the usefulness of the developed models and the production rules. Objective has been to find out whether the discussed way facilitates integration of two models to one virtual prototype. Therefore, a Capture-the-Flag (CtF) application has been used. Originally CtF is a game where a hunter needs to capture a flag, the other players chase the hunter and try to prevent him from capturing the flag. In our case the players are the BeBots; one is the hunter, the other try to chase it. The BeBots operate autonomously without any interaction by a user. Figure 4 shows two screenshots of the application. The left figure shows an overview of the virtual environment. The flag stands in the middle of the environment. Spheres are placed as obstacles. The BeBots need to avoid them. The right figure shows a detailed view to the scene. The BeBot with the yellow diamond on top is the hunter and tries to capture the flag.

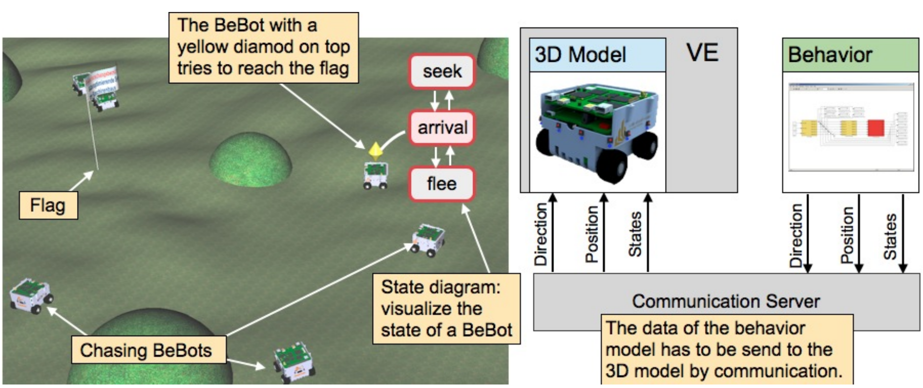


Fig. 4. Overview of the application (left), the architecture of the application (right)

Each robot is represented by a 3D model and a behavior model. Both models have been annotated by the RDF notation [16]. The annotation of the 3D model describes the input variables to set the position and direction of a robot as well as a state diagram to visualize its current state. The behavior model provides the position and direction of each robot. Both applications (VE and behavior) need to be linked by the agent respectively the agent has to identify the variables and to link them.

In summary the agent has been able to realize the communication between both models / applications utilizing the RDF-based annotations of both models. The desired application could be realized, without any need for a user to describe the communication manually.

5 Conclusion and Outlook

This paper introduces software agents that facilitate the automatic integration of different aspects model of virtual prototypes. The aspect models become annotated by and RDF-notation. The notations of two models are processed and a vector is determined that expresses the similarity of two models. The concept has been tested. Therefore, an application has been developed that includes different mobile robots. Task of the agent has been to integrate the models into a template of a virtual prototype. The results show us that the concept can be utilized for the desired task. The results are a first step, but two things could be shown by this work:

First, a RDF description for the different aspect models is the right way to annotate the different models. The resources and properties of the developed RDF notation facilitate the computer-internal annotation of the task, the visualization, and the functions. They can be used as meta data for these aspects.

Second, software agents are a suitable solution to structure the software for the desired task. The agent communication allows the loading of different RDF notation. Furthermore, production rules can be utilized as internal agent model, as knowledge base, that drives the decision.

In summary the approach works well, when two models are used that represent different aspects (shape, behavior) of an equal technical system like the BeBot.

Altogether the results are a good starting point for further research. Until now, the carried out tests show us the usefulness of the concept at separated applications. The concept has not been generalized. In order to generalize it, a common set of production rules needs to be identified. For this purpose, we will perform a study utilizing different application examples with different VPs. On this basis a common set of production rules will be developed.

Acknowledgement. This contribution was developed and published in the course of the Collaborative Research Center 614 “Self-Optimizing Concepts and Structures in Mechanical Engineering” funded by the German Research Foundation (DFG) under grant number SFB 614.

References

1. Krause, F.-L., Jansen, C., Kind, C., Rothenburg, U.: Virtual Product Development as an Engine for Innovation. In: Krause, F.-L. (ed.) Proceedings of the 17th CIRP Design Conference on The Future of Product Development. Springer, Berlin (2007)

2. Gausemeier, J., Ebbesmeyer, P., Kallmeyer, F.: Produktinnovation. Carl Hanser Verlag, München (2001)
3. Gutierrez, M., Vexo, F., Thalmann, D.: Stepping into Virtual Reality. Springer, Berlin (2007)
4. Mendez, G., de Antonio, A.: An Agent-Based Architecture for Collaborative Virtual Environments for Training. In: Proceedings of the 5th WSEAS Int. Conf. on Multimedia, Internet and Video Technologies, Corfu, Greece, August 17-19, pp. 29–34 (2005)
5. Galea, A., Borg, J., Grech, A., Farrugia, P.: Towards Intelligent Design Tools for Micro-scale components. In: International Conference on Engineering Design, ICED 2009, Stanford, CA, August 24-27, pp. 5-73 – 5-84 (2009)
6. Payne, T.R.: Agent-based Team Aiding in a Time Critical Task. In: HICSS 2000, Proceeding of the 44th Hawaii International Conference on System Sciences, vol. 1 (2000)
7. Baolu, G., Shibo, X., Meili, C.: Research and Application of a Product Cooperative Design System Based on Multi-Agent. In: Third International Symposium on Intelligent Information Technology Application, pp. 198–201 (2009)
8. Geiger, C., Lehrenfeld, G., Mueller, W.: Authoring communicating agents in virtual environments. In: Proceedings of the Computer Human Interaction, Adelaide, SA, Australia, pp. 22–29 (1998)
9. Bludau, C., Welp, E.: Semantic Web Services for the Knowledge-based Design of Mechatronic Systems. In: Proceedings of the International Conference on Engineering Design, ICED 2005, Melbourne, Australia, August 15-18 (2005)
10. Restrepo, J.: A Visual Lexicon to Handle Semantic Similarity in Design Precedents. In: Proc. of the 16th International Conference on Engineering Design, ICED 2007, Paris (2007)
11. Ding, L., Matthews, J., Mullineux, G.: Annacon: Annotation with constrains to support design. In: International Conference on Engineering Design, ICED 2009, Stanford, CA, August 24-27, pp. 5-37 – 5-48 (2009)
12. Ding, L., Davies, D., McMahon, C.A.: The integration of lightweight representation and annotation for collaborative design representation. Research in Engineering Design 19(4), 223–238 (2009)
13. Li, C., McMahon, C., Newnes, L.: Annotation in Design Processes: Classification of Approcches. In: International Conference on Engineering Design, ICED 2009, Stanford, CA, August 24-27, pp. 8-251 – 8-262 (2009)
14. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American (2001)
15. Pahl, G., Beitz, W., Feldhusen, J., Grote, K.-H.: Engineering Design: A Systematic Approach, 3rd edn. Springer, Berlin (2006)
16. Radkowski, R.: Towards Semantic Virtual Prototypes for Automatic Model Combination. In: 20th CIRP Design Conference, Global Product Development, Nantes, France (2010)
17. Radkowski, R., Waßmann, H.: Software-Agent Supported Virtual Experimental Environment for Virtual Prototypes of Mechatronic Systems. In: Proceedings of the ASME 2010 World Conference on Innovative Virtual Reality WINVR 2010, Ames, Iowa, USA, May 12-14 (2010)

Virtual Factory Manager

Marco Sacco[1], Giovanni Dal Maso[2], Ferdinando Milella[3], Paolo Pedrazzoli[4], Diego Rovere[4], and Walter Terkaj[1]

[1] ITIA-CNR, Institute of Industrial Technologies and Automation, National Research Council, via Bassini, 15 - 20133 Milano, Italy
{marco.sacco,walter.terkaj}@itia.cnr.it
[2] Technology Transfer System S.r.l., via Pacini 15, 20131 Milano, Italy
dalmaso@ttsnetwork.com
[3] SimX Ltd., Furness House, Salford Quays, M50 3XA, Manchester, UK
f.milella@simx.co.uk
[4] ICIMSI, Institute of Computer Integrated Manufacturing for Sustainable Innovation, Galleria 2, Manno, Switzerland
{paolo.pedrazzoli,diego.rovere}@icimsi.ch

Abstract. The current challenges in manufacturing engineering are the integration of the product/process/factory worlds (data and tools) and the synchronization of their lifecycles. Major ICT players already offer all-comprehensive Product Lifecycle Management suites supporting most of the processes. However, they do not offer all the required functionalities and they lack of interoperability. An answer will be given by the development of a Virtual Factory Framework (VFF): an integrated virtual environment that supports factory processes along all the phases of its lifecycle. This paper will focus on the Virtual Factory Manager (VFM) that acts as a server supporting the I/O communications within the framework for the software tools needing to access its data repository. The VFM will ensure data consistency and avoid data loss or corruption while different modules access/modify partial areas of the data repository at different times. Finally, an industrial case study will show the potentiality of the VFM.

Keywords: Virtual Factory, Interoperability, Reference Model.

1 Introduction

Manufacturing has to cope with a complex and evolving market environment. While the world crisis breaks the balance between demand and production, the global market pushes for a continuous change. Several critical aspects related to the rapid prototyping of factories have to be addressed. It is critical to provide sufficient product variety to meet customer requirements, business needs and technical advancements [1], while maintaining economies of scale and scope within the manufacturing processes [2]. Therefore, the current challenge in manufacturing engineering consists in the innovative integration of the product, process and factory worlds and the related data, aiming at synchronizing their lifecycles [3]. This synchronization can be empowered by digital technologies, as it is shown both by

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 397–406, 2011.

industrial practice and academic scientific research. Indeed, the topic of Digital and Virtual Factory [4, 5, 6, 7] have been addressed by several research projects, such as METNET, SPAS, MANUVAR, EMI, EMERALS [8], VFTS [9], IRMA, DiFac [10] and Dorothy [11]. The Virtual Factory (VF) paradigm can assist a production environment by addressing various key issues like: (1) reduction of production times and material waste thanks to the analysis of virtual mock-ups, (2) development of a knowledge repository where people can find stored information in different versions, with both advisory role and support to the generation of new knowledge, (3) improvement of workers efficiency and safety through training and learning on virtual production systems, (4) creation of a collaboration network among people concurrently working on the same project in different places.

The complexity of the problem calls for support tools to effectively address all the phases of the factory lifecycle. Indeed, the ICT players (e.g. Siemens PLM, PTC and Dassault Systèmes) already offer all-comprehensive suites containing software tools that have been developed or acquired in the recent years. These tools deal with most of the factory planning, design and deployment phases. However, the current approaches still do not meet the demands of the industry and fail to provide all the required functionalities. One of the reason is the lack of interoperability. Moreover, Small and Medium Enterprises cannot afford the present expensive PLM software suites.

The previous analyses highlight the need of a novel framework for the VF enabling a step forward in the state of the art, by describing the factory as a whole consisting of processes, dependencies and interrelations, data and material flows [12]. This framework should guarantee the democratization of industrial systems for simulation and analysis, by reducing the total cost of ownership of a holistic virtual representation of the factory. In particular, the following aspects of democratization need to be properly addressed:

1. *Decrease of the investment and operating costs* that are currently associated with the commercial all-comprehensive software suites.
2. *Wider range of users*. The functionalities and usage of simulation and virtual factory technologies should be extended from the workstations/desktops of the engineers to the laptops of managers and marketing executives.
3. *Usage and management*. Democratization also means giving engineers and field technicians the ability to take advantage of simulation and virtualization of processes without relying on dedicated specialists.

This paper presents a framework for the VF that is carried out by the European project “Virtual Factory Framework” [13]. The framework is introduced in Section 2, whereas Section 3 focuses on the development of its main software components. Finally, Section 4 shows the results obtained with the first implemented prototypes.

2 Virtual Factory Framework

The Virtual Factory Framework (VFF) can be defined as “An integrated collaborative virtual environment aimed at facilitating the sharing of resources, manufacturing information and knowledge, while supporting the design and management of all the factory entities, from a single product to networks of companies, along all the phases of the their lifecycles". The VFF architecture (see Fig. 1) is based on four Pillars: (I) *Reference Model*, (II) *Virtual Factory Manager*, (III) *Decoupled Functional Modules* and (IV) *Integration of Knowledge*. The key characteristics of the pillars are openness, scalability and easiness to plugin the decoupled software tools, thus reducing the *investment costs* compared to "all-in-one" software suites. Moreover, the VFF aims at promoting major *time and operating cost* savings, while increasing the performance in the design, management, evaluation and reconfiguration of new or existing factories.

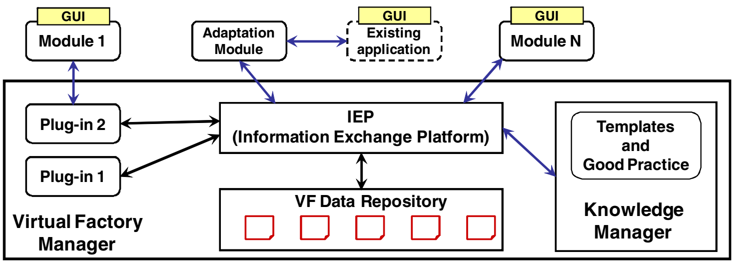


Fig. 1. Virtual Factory Framework architecture

The *Reference Model* establishes a coherent standard extensible Virtual Factory Data Model for the common representation of factory objects related to production systems, resources, processes and products. The common data model can be considered as the shared meta-language providing a common definition of the data that will be governed by the Virtual Factory Manager (Pillar II) and used and updated by the Decoupled Functional Modules (Pillar III). Since the VFF has to deal with the interactions between several activities, it is necessary to take care of data evolution along the factory lifecycle phases as well.

The *Virtual Factory Manager* (VFM) is the core of the VFF and handles the common space of abstract objects representing the factory as defined by the common data model (Pillar I). The VFM orchestrates the decoupled functional modules and provides a controlled access to the different virtual factory instances. Considering the characteristics of the data model and the need to interface several decoupled modules, the structure of the VFM has been designed by adopting *star network* architecture as shown in Fig. 1. Next section will delve into the structure of the VFM by presenting its main characteristics and the development of the first prototype.

The *Decoupled Functional Modules* (named *VF modules*) are the software tools that implement the various methods and services to support the activities related to factory design, performance evaluation, management, production monitoring, etc. The VF modules can be located on a remote workstation or on the server where the VFM resides. Considering the scope of the VFF approach, the VF modules can be grouped into categories. For each category, different solutions can be adopted according to the specific needs and the availability of commercial applications. The integration of VF modules endowed with different functionalities and level of detail but insisting on the same factory representation will offer the possibility to reach a *wide range of users*.

Integration of Knowledge aims at supporting the modeling of complex systems and providing greater comprehension of the business processes. The Knowledge Manager (see Fig. 1) is responsible of the knowledge repository that is designed according to an ontology-based approach. A knowledge association engine will extract the knowledge from the repository by means of rule-based mechanisms and case-based reasoning techniques. The exploitation of embedded and constantly growing knowledge presented as good practice and templates (i.e. toolsets of pre-dealt problems) will enable the engineers to *easily use and manage* the virtual factory without specific expertise or experience.

3 Virtual Factory Manager

This section presents the analysis of the requirements for the VFM (Sect. 3.1) and its proposed architecture (Sect. 3.2). Finally, the VFM prototype is described (Sect. 3.3).

3.1 VFM Requirements

The main goal of the VFM design and implementation consists in obtaining an open integration platform representing a common and shared communication layer between already existing and newly developed software tool to support the factory design and management. This final goal leads to the definition of several requirements:

- *Platform independent interfacing capabilities*. The VF modules are software tools developed by different vendors/organizations, with different programming languages, operating systems and HW architectures. The VFM has to interface all of them by providing its services in an open and "proper" way.
- *Management of concurrent access and data consistency.* Several software tools can access and/or modify partial areas of the factory data at different, and possibly overlapping, times. Therefore, the VFM is required to ensure that concurrent accesses occur without endangering the data integrity and slowing down the planning process to unacceptable levels.
- *Management of evolving Factory Data*. The VFM has to provide functionalities for managing the evolution and revision of the data related to complex entities like production systems, processes and products.
- *Data safety* must be ensured in case of hardware failures or user errors.
- *Addition of customized functionalities.* Third party developers need an appropriate mechanism to enrich the set of functionalities provided by the VFM without impacting on its core.
- *Response time*. The interaction between the VFM and the VF modules requires the support of communication mechanisms that are able to provide answers in an appropriate time frame.

3.2 VFM Architecture

The architecture of the VFM was designed to provide support to the required functionalities. Each solution implemented by the VFM is based on stable and well-established technologies in order to obtain an overall system capable to respond to industrial needs of reliability. The resulting VFM architecture is shown in Fig. 2 as an UML component diagram.

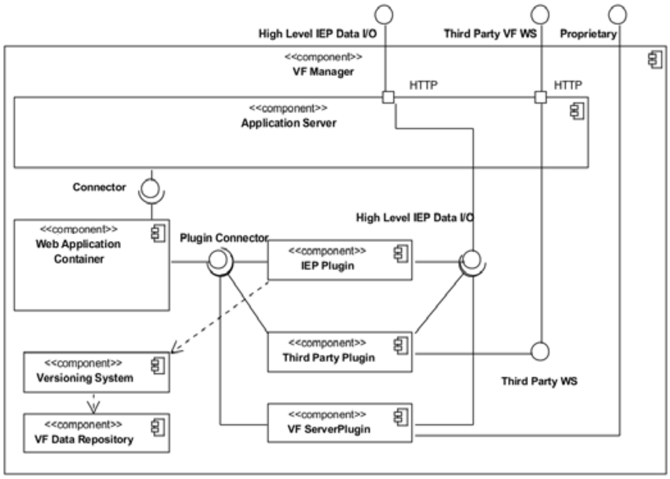


Fig. 2. UML component diagram of the VFM

The *VF Data Repository* is the central data repository where all the shared data will be stored. The *evolution of the factory data* is managed by the *Versioning System* that organizes and updates the set of virtual factory instances. The Versioning System guarantees the *data safety* as well, since it allows restoring an older version at anytime, thus preventing data losses due to user errors. Moreover, rollback methods can be used in case of data inconsistencies due to broken connections or other factors, always ensuring data safety.

The functionalities of the VFM are exposed as web services that have been identified as a suitable and widely adopted solution to *guarantee platform independent interfacing capabilities*. The *Application Server* provides the front end for the exposure of VFM functionalities and takes care for the information transport of the VFM. The *Web Application Container* is the component of the architecture that provides a platform for implementing and publishing VFM functionalities.

The *Information Exchanging Platform (IEP)* is the main component of the VFM and provides VF modules and *plugins* with a high level access to the data stored in the VFM. It represents the preferred (even if not the only one) way to connect to the VFM, since it provides a complete set of methods for structured data retrieval and validation, data locking mechanism and factory version management. In particular, the locking mechanism helps to manage the *concurrent access* of the VF modules.

A *Plugin* is a component that enables the VFM to be scalable by hosting server side third party software, thus *adding customized functionalities*. Server side software packages provide their own specific services interfacing the containing framework that ensures local optimized access to the *IEP* and support to a container already configured for exposing methods as web services.

The interaction between the VFM and the VF modules mainly consists in an exchange of data streams only after specific requests (e.g. download/upload of data, locking of data, etc.) are made by the VF module. This kind of interaction requires no "real-time" *response* from the VFM. Nevertheless, if some module has specific requirements on the type of interaction (e.g. for response time or for data encryption), the implementation of a server side plugin with a dedicated proprietary interface is always possible and supported by the VFM structure.

3.3 VFM Prototype

The implementation process of the VFM prototype was driven by the adoption of open source platforms while following the architecture in Fig. 2. This decision was taken to obtain a completely open solution that can be developed and maintained by different actors. Java was chosen as software platform because of the availability of high-level and reliable tools that enable the application in real industrial scenarios.

The *VF Data Repository* was developed as a file system where all the data are stored on the server in form of files. The adoption of a file-based system instead of a Database Management System (DBMS) is justified by its flexibility and the possibility to apply a versioning system. Most of the data are stored in XML files [14] that can be validated by a set of XSD (XML Schema Definition) [15] files representing the the VF Data Model. Besides XML files, the file system can host files of any kind (e.g. binary files for graphical and geometric representations).

The *Versioning System* was developed by adopting Subversion [16] that is a widespread open source version control system. Its original purpose is to maintain versions of software source code and text documents, but it can be used to track changes in any kind of file and it is suitable for the VFM needs since it is efficient when applied to text-based files as the XML files are.

The *Application Server* was implemented as an Apache HTTP Server [17] that is an open source modular web server and one of the most deployed HTTP servers. Moreover, Apache is well known for being reliable and for supporting most of the programming languages and platforms.

The *Web Application Container* was developed using Apache Tomcat [18] that is an open source project of the Apache Software Foundation. It powers several large-scale and mission-critical web applications for a wide range of companies and organizations. Tomcat can be paired with "Tomcat mod" that is a supporting component required to forward the information received by the Apache Server to Tomcat, and then to the plugins.

The *IEP* component was implemented as a Tomcat web application and its functionalities are exposed as a set of *cross-platform* web services based on SOAP (Simple Object Access Protocol) [19], thus enabling any VF module to use them. The IEP prototype provides both automated versioning and locking mechanisms to prevent data inconsistency thanks to "check-out" and "commit" operations.

4 Testing the VFM Prototype

This section explores the potentiality of the VFF and in particular of the VFM by showing how different software tools can interoperate while addressing the same industrial problem. In particular, the test case is focused on the design of a factory layout for a Romanian company (Compa S.A.) playing in the automotive market. A reduced version of the final VF Data Model was developed as an XSD file and three software tools were deployed in the VFF as VF modules: *GIOVE Virtual Factory* [20], *Factory Layout Planner* [21], and *3DCreate* by *Visual Components Oy* [22].

GIOVE Virtual Factory (*GIOVE VF*) is a virtual reality collaborative environment aimed at supporting the factory layout design. Machines, operators and other resources can be selected from a library and placed in the 3D scene of the virtual factory. The virtual environment can schematically display performance measures that are provided by simulators and/or monitoring tools. Thanks to its user-friendly interface, GIOVE VF enables the collaboration between managers, experts and also workers in an intuitive way.

Factory Layout Planner (*FLP*) is a client/server application that enables the collaborative development of a factory layout thanks to three key features: the 3D visual editing of the layout, the possibility to act on the same layout in a distributed environment, the ability to perform Discrete Events Simulation (DES) on the layout. The collaboration on the layout can be both remote and local; the former allows user distributed all over the world to cooperate in the layout creation, the latter allows users to act on the same device at the same time on a common model. This functionality can be achieved thanks to the integration of multi-touch tables.

3DCreate by *Visual Components Oy* performs material flow and robotic simulation on the same platform for simulation and visualization of complete manufacturing systems. 3D equipment models can be created with increasing level of details to a point where they represent real factory counterparts in look and behaviors. New factory layouts are created from a catalog by simply snapping equipment models together. The user can run simulations and perform a large number of general layout validations, like collision detection, resource utilizations, cycle times, etc.

The interoperation between the previously described software tools represents an interesting test bed for validating the VFM concept because:

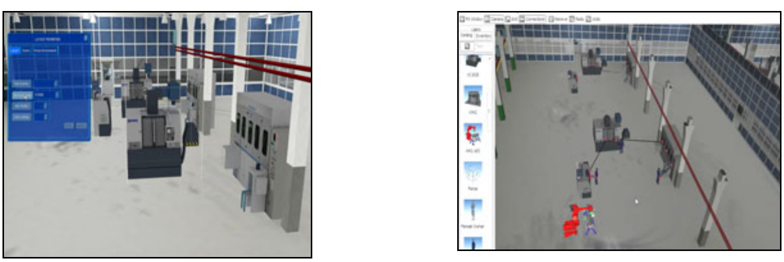
- The tools were developed with different programming language and can operate on different OS and platforms, thus enabling the validation of the VFM "universality": i.e. GIOVE VF was developed in C++, whereas FLP in Java. Visual Components, developed in C++, was connected to the VFM through Python.
- The tools share some functionalities (e.g. visualization of a 3D factory layout), thus enabling a validation of the VFM functionalities.
- The tools have complementary functionalities (e.g. navigation in a 3D environment, DES and kinematic simulation provided by GIOVE VF, FLP and Visual Components, respectively), thus showing the benefits of interoperability.

During the test each software tool had a different user and was run on a different computer. The three software tools were connected to a server hosting the VFM and exposing its web services. The basic VFM functionalities were successfully tested since all the three VF modules could access the web services exposed by the VFM, thus meeting the requirement of *platform independent interfacing capabilities*. The XML files in the VF Data Repository were correctly imported and interpreted by the three tools thanks to the common data model. Figure 3 shows how the three tools offer only a slightly different graphical representation of the same factory view.

(a) (b) (c)

Fig. 3. Same factory view offered by (a) GIOVE VF, (b) FLP and (c) Visual Components

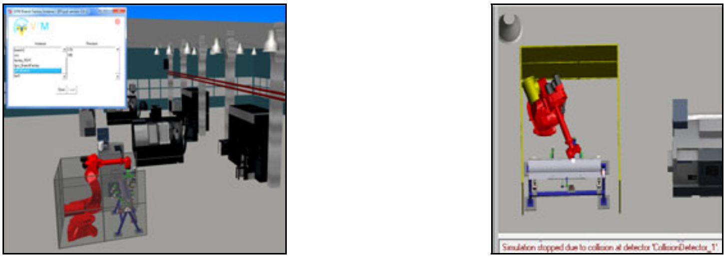
The interoperation between the VF modules was tested by incrementally modifying a factory layout in a collaborative way, starting from a simple layout and then gradually increasing its complexity by adding production resources and using functionalities provided by the VF modules. Every step was performed through a check-out/modify/commit cycle. The VF modules have to access the factory data in a selective way, leaving untouched data that the VF module does not use and/or is not capable of using. For example, after the GIOVE VF user created a draft factory layout (Fig. 4.a), the FLP user modified the layout by adding new machines and operators and creating connections between the resources (Fig. 4.b), so that a DES could be run afterwards. The modifications implemented by the FLP user were finally checked and correctly visualized by the GIOVE VF user, thus demonstrating how the VFM is able to meet the requirement of guaranteeing the *data consistency* across VF modules.



(a) (b)

Fig. 4. (a) The GIOVE VF user and (b) the FLP user modify the same factory layout

Already existing factory instances can be used as a starting point to develop new layouts or when a stakeholder wants to evaluate "what if" scenarios without overwriting the existing data and upsetting the lifeline of a factory instance. A new instance can be created from an existing one at any stage of its lifeline thanks to the "branch-out" functionality provided by the IEP. For example, the Visual Components user created a new factory instance starting from the version previously saved by the GIOVE VF user, thus showing how the VFM can *manage evolving factory data*. Then the Visual Components user modified the layout by adding a fence around the robot (Fig. 5.a) and checked potential collisions of the robot by means of a kinematics simulation (Fig. 5.b). Finally, the *management of concurrent access* was tested too. If a VF module tries to access a factory version that is currently locked by another user, then the request is denied and the factory data can be downloaded in read-only mode.



(a) (b)

Fig. 5. A Visual Components user (a) adds a fence to the layout and then (b) performs a kinematics simulation of the robot to detect possible collisions with the fence

5 Conclusions

The proposed framework and the VFM provide a concrete answer to the requirements of interoperability to address the product/process/factory lifecycle. The VFM prototype and the feasibility studies presented in this paper represent a first cornerstone for future developments.

The use of XSD/XML has guaranteed to have a pre-defined syntactic structure for the data (XML) that can be validated according the XSD definitions. Moreover, XSD/XML represents an expressive technology where several default data-types can be extended and complex constraints and properties can be modeled. However, the complex processes associated with the virtual factory require also the support of knowledge and the characterization of data with their relations on a semantic level. This is hardly given by the current data representation based on XSD, since inter-document references (cross-references) are poorly modeled, thus endangering the referential consistency and the management of distributed data. Therefore, the possibility to adopt an ontology-based representation of data and exploit the Semantic Web technologies for the VFM will be evaluated.

In the coming months more complex industrial cases will be addressed and more software tools will be integrated to present a complete solution for the whole factory lifecycle. Finally, further functionalities will be added to the VFM.

Acknowledgments. The research reported in this paper has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No: NMP2 2010-228595, Virtual Factory Framework (VFF). The authors would like to thank COMPA S.A. (Sibiu, Romania) for kindly providing the industrial case information.

References

1. Huang, G.Q., Simpson, T.W., Pine II, B.J.: The power of product platforms in mass customization. International Journal of Mass Customisation 1(1), 1–13 (2005)
2. Terkaj, W., Tolio, T., Valente, A.: Designing Manufacturing Flexibility in Dynamic Production Contexts. In: Tolio, T. (ed.) Design of Flexible Production Systems, pp. 1–18. Springer, Heidelberg (2009)
3. Tolio, T., Ceglarek, D., ElMaraghy, H.A., Fischer, A., Hu, S., Laperrière, L., Newman, S., Váncza, J.: SPECIES – Co-evolution of Products, Processes and Production Systems. CIRP Annals - Manufacturing Technology 59(2), 672–693 (2010)
4. Zöllner, M., Keil, J., Behr, J., Gillich, J., Gläser, S., Schöls, E.: Coperion 3D – A Virtual Factory on the Tabletop. In: Proceeding of 5th INTUITION International Conference: Virtual Reality in Industry and Society (2008)
5. Ding, J., Wang, Y., Chen, K.: An Interactive Layout and Simulation system of Virtual Factory. Applied Mechanics and Materials 20-23, 421–426 (2010)
6. Yang, S., Ahn, B., Seo, K.: Development of a prototype customer-oriented virtual factory system. The International Journal of Advance Manufacturing Technology 28(9-10), 1031–1037 (2006)
7. Zhai, W., Fan, X., Yan, J., Zhu, P.: An Integrated Simulation Method to Support Virtual Factory Engineering. International Journal of CAD/CAM 2(1), 39–44 (2002)
8. Jain, S., Choong, N.F., Aye, K.M., Luo, M.: Virtual factory: an integrated approach to manufacturing systems modeling. International Journal of Operations & Production Management 21(5/6), 594–608 (2001)
9. Kazlauskas, E.J., Boyd, E.F., Dessouky, M.M.: The Virtual Factory Teaching System (VFTS): Project Review and Results. In: Proceedings of ED-MEDIA 2002 World Conference on Educational Multimedia, Hypermedia & Telecommunications (2002)
10. Sacco, M., Redaelli, C., Cândea, C., Georgescu, A.V.: DiFac: an integrated scenario for the Digital Factory. In: Proceedings of 15th International Conference on Concurrent Enterprising (2009)
11. Pedrazzoli, P.: Design Of customeRdRivenshOes and multisite factory – DOROTHY. In: Proceedings of 15th International Conference on Concurrent Enterprising (2009)
12. Pedrazzoli, P., Sacco, M., Jönsson, A., Boër, C.: Virtual Factory Framework: Key Enabler For Future Manufacturing. In: Cunha, P.F., Maropoulos, P.G. (eds.) Digital Enterprise Technology, pp. 83–90. Springer, Heidelberg (2007)
13. Sacco, M., Pedrazzoli, P., Terkaj, W.: VFF: Virtual Factory Framework. In: Proceedings of 16th International Conference on Concurrent Enterprising, Lugano, Switzerland (2010)
14. Extensible Markup Language (XML), http://www.w3.org/XML/
15. XML Schema Part 1: Structures 2nd edn, http://www.w3.org/TR/xmlschema-1/
16. Collins-Sussman, B., Fitzpatrick, B., Pilato, M.: Version Control with Subversion (2008), http://svnbook.red-bean.com/index.en.html
17. Apache, HTTP Server Project, http://httpd.apache.org/
18. Apache Tomcat 6.0, http://tomcat.apache.org/tomcat-6.0-doc/index.html
19. SOAP Specifications, http://www.w3.org/TR/soap/
20. Viganò, G.P., Greci, L., Sacco, M.: GIOVE Virtual Factory: the digital factory for human oriented production systems. In: Proceedings of the 3rd International CARV Conference, Munich, Germany, pp. 748–757 (2009)
21. Ceruti, I.F., Dal Maso, G., Ghielmini, G., Pedrazzoli, P., Rovere, D.: Factory Layout Planner. In: Proceedings of 16th International Conference on Concurrent Enterprising, Lugano, Switzerland (2010)
22. 3DCREATE Visual Components, http://www.visualcomponents.com/Products/3DCreate

FiveStar: Ultra-Realistic Space Experience System

Masahiro Urano[1], Yasushi Ikei[1], Koichi Hirota[2], and Tomohiro Amemiya[3]

[1] Tokyo Metropolitan University
6-6, Asahigaoka, Hino-shi, Tokyo 191-0065, Japan
ikei@computer.org, urano@krmgiks5.tmit.ac.jp
[2] Tokyo University
5-1-5, Kashiwanoha, Kashiwa-shi, Chiba 277-8563, Japan
hirota@media.k.u-tokyo.ac.jp
[3] Nippon Telegraph and Telephone Corporation
3-1, Morinomiyawakamiya, Atugi-shi, Kanagawa 243-0198, Japan
amemiya.tomohiro@lab.ntt.co.jp

Abstract. This paper describes the development of the FiveStar system that provides five senses stimulations to the participant for the creation of ultra-realistic experiences. We performed an upgraded demonstration of the system to evaluate its individual technologies at Asiagraph 2010 in Tokyo. The content of the exhibit was the encounter with a yokai character that produces effects of extra-ordinary interaction between the participant and the imaginary characters. The experiences of participants were investigated as exploratory effort for the elucidation of this type of ultra-reality created in a fantasy world.

Keywords: Multiple-modality, Interactive Experience, Augmented Reality, Ultra Reality.

1 Introduction

Recently, a three-dimensional (3D) display has attracted people's attention since many manufacturers announced that a stereoscopic TV set will be distributed worldwide in the year of 2010. Several 3D movies have already been released from the movie makers that intend to shift the 3D technology from a theme park to theaters in a city and then to a living room. The 3D visual experience provided in such a movie is much more sophisticated as a whole than those created in the past. It gives more augmented sensation of presence than 2D movies with improved high-density digital image generation. However, the presence we get in the real environment is not created only with visual information but with all the modalities that make our lives truly real.

A new project on the development of a multimodal display has started in 2009 under the administration of Tokyo Metropolitan University. The project aims to develop a new display system that presents ultra realistic sensation by providing cues to multiple modalities [1]. The display system is named ‘Five Senses Theater’ (FiveStar) since it is intended to produce controlled stimuli to all human sensors, the five senses except for the gustatory sensation. The FiveStar is an interactive display system for a single user of home theater size. A three-dimensional visual display, a spatial sound, a haptic/tactile display, a wind and scent display, and a vestibular display are involved in the FiveStar that is targeted to produce a futuristic integrated

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 407–414, 2011.

ultra-reality [2] experience. To develop each component of the system properly, integration testbed contents were built first based on individual function arrangement with intuitive concept design.

2 Five Senses Theater

2.1 Basic Configuration

Figure 1 shows a design sketch of the Five Senses Theater submitted as a proposal at the beginning of the project. This illustrates hardware elements that form the system with their input data and output targeted sensations. Users get multiple-modality stimulations simultaneously from the system. The system consists of a stereoscopic visual display, a 7.1 channel surround audio system, a tactile and force feedback display system, a wind and scent delivery system, and a vestibular body motion system. These subsystems are integrated to perform an interaction with a participant along with a particular experience scenario the participant immerses in. The system is intended for a personal use hypermedia unit in the near future.

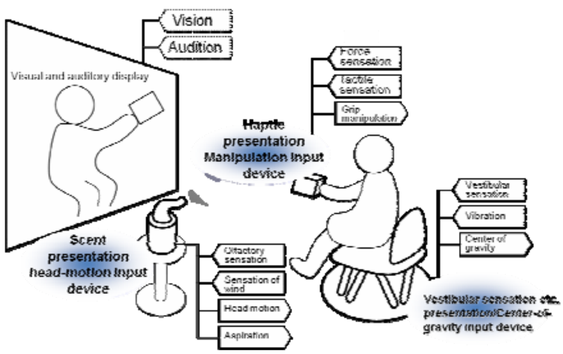


Fig. 1. First design sketch of Five Senses Theater. This is intended to present a particular kind of ultra reality. Multiple-modality sensations are integrated to render alternative reality and communication.

More specifically, the objective of the project is the development of core technology to build a platform for the five-sensory communication and five-sensory content experience. It aims to augment a uni-modal display device to an ultra-reality media with five-sensory presentation and interaction used in the communication. The target also includes the augmentation of rendering of experience of movies and video games.

For the creation of the ultra reality on the display device, we focus more on psychological effects than the fidelity of each device that has often technical limitation at present relative to the quality of human senses. Thus, what is to be pursued here is the effective rendering for impressive experience in which the impressiveness and richness of experience are the principal points of evaluation. Therefore, the basic idea is that the ultra-reality is not the extension of high fidelity-oriented reality. Of course the quality of presentation of each system is the bottom line. However, the integration of presentations in modalities to the user's intention is the point of the system.

2.2 Conceptual Level of Development

To tackle with the ultra-reality, we assumed a multilevel structure for the FiveStar design. Figure 2 shows the conceptual levels that have three levels, Scene level, Sensation/Manipulation level and Device level. Actual interactions are exchanged at the Device level between sensors and motors of both sides of the user and the FiveStar. In this bottom layer, only limited information can be exchanged due to the limited capability of the physical device. The information at the level is described based on physical quantity.

The second level of Sensation and Manipulation level determines the range of device level to effectively produce the interaction performed by the user. The function of the second level is described by sensory information and motion input. The second level is designed to implement the upper level functions. The top level of Scene level handles the context to be exchanged through the FiveStar by the user. The scene is described at the cognitive and emotional points of views. We assume that the scene is rendered to best express or transfer the intention of the user. The five senses need to be integrated in this direction.

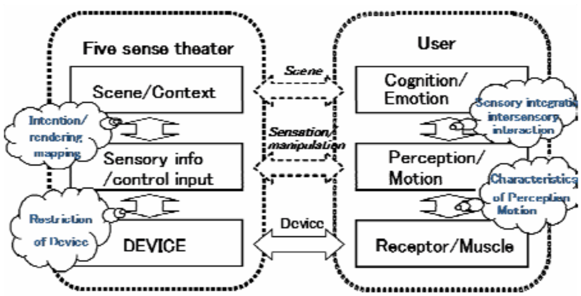


Fig. 2. Conceptual levels of Five Senses Theater design principle. Devices are only used to fulfill cognitive/emotional requirements placed essentially by the user.

2.3 Challenges and Trials

Challenges are how effectively transmit the impressive and eloquent information based on the multiple-modality interface devices. The presentation needs to be optimized on the cognition and interpretation level. For this purpose, we consider that stimulation is presented rather approximately than physically precisely. Also the input motion is interpreted rather approximately than investigated very precisely. The coherence of the objectives and the sensorimotor information exchange should be maintained in the interaction design. A support mechanism might be installed to control consistency.

For the embodiment of ultra-reality, we started investigation on the techniques to transmit reality more effectively than the usual approach that rebuilds the reality as it looks in the real world. The ultra-reality itself is an unestablished idea that will welcome an original proposal. We consider the ultra-reality has many aspects. Of course, it is more realistic than ever in a sense. But it is not necessarily higher definition reality than ever by higher resolution display. Probably, high resolution will not be solely sufficient for us. It is rather an issue of multiple-modality integration and rendition through the user's intention.

3 Prototype Exhibit at Asiagragh 2010

The development of the each unit of every modality was focused so that the system could render a specific scene that worked as a bench mark of the whole system. The scenes were tested in technology exhibits where the prototype of FiveStar was first demonstrated in 2009 as an early stage evaluation of its element technology. Then, the prototype was improved and presented again as part of Asiagraph 2010[1] in Tokyo. The original exhibit was named 'Ultra-ordinary experience through five senses media.' The content of the exhibit was an experience of a fantasy space: a participant who sat on the system was brown to an another dimension where a specter lived and casted evil magic particles to the participant. Then the system produced a multiple-modality interaction to the participant. This is one of approaches to embody and investigate how we think the ultra reality is. The reality has a number of aspects to be expressed differently by a media. The content of this exhibit was a trial to render the reality at the view of an imaginary world.

3.1 System Configuration

Figure 3 shows the configuration of the system. The system consists of a haptic (tactile and force) system to present physical reaction of a particular event, a stereoscopic visual system to render the world and to show the image of a participant using a stereo camera. A 7.1 surround audio, an eight-channel wind scent source system, a compressed air jet system, tactile stimulators to the body on the chair, a 3D motion seat, and a 2D feet motion generator system were also installed in the system to create integrated stimuli to the participant. Figure 4 shows the FiveStar system overview at the exhibit. The system was assembled at the site of the ARTEC exhibit in Asiagraph.

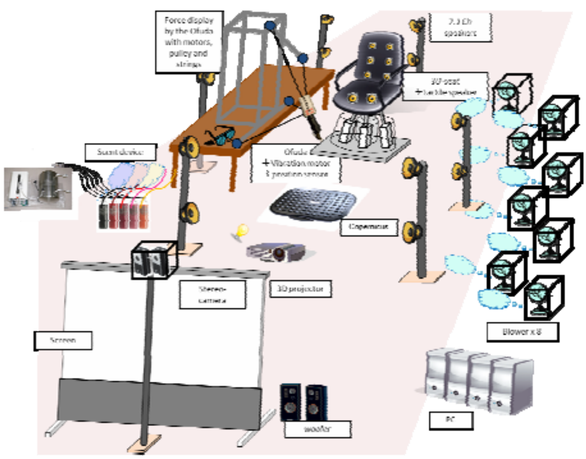


Fig. 3. System configuration of FiveStar exhibit in 2010

[1] National Museum of emerging science and innovation in Tokyo.

3.2 Theme of Exhibit

The theme of demonstration we created was a short-time encounter experience with the 'Yokai' which is a Japanese traditional specter. The scenario of this experience consists of three parts. First, the participant was brown by a wind caused by a fanciful creature from the site to an another world. Then evil specter appeared casting the magic particles from its mouth. The participant protects the particles of magic by holding a shield with an amulet called 'ofuda'. After the participant received the magic particles sufficiently into the amulet, the participant could throw the amulet to the specter. When the amulet defeated the specter, it changed to a friendly character. Then the specter let the participant be back to the original world where the participant was sitting. Thus, a participant was able to experience and interact with an extraordinary world with multiple sensations.

4 Displays of the System

We built a personal interactive theater where the response from the simulated world was designed best to fit to the participant's imagination of the story. This is a working prototype that we assume a personal use hypermedia unit for the future. Figure 4 shows the system setup built at Asiagraph 2010.

4.1 Haptic System for the Amulet (ofuda)

The haptic system consists of a 3D force-feedback device which is called 'elePhantom' and a six-channel tactile device ('ofuda' device). The haptic system is shown in Figure 5. A participant held the ofuda device that presented reaction force by the elePhantom at the hand and tactile sensation on the palm during the interaction with the VR space.

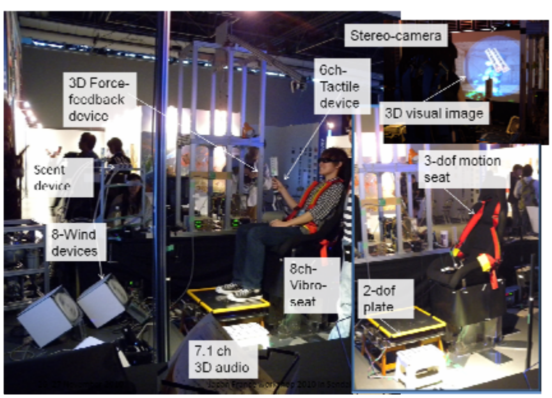


Fig. 4. System overview of FiveStar exhibit in 2010

The force feedback device (elePhantom) consists of four motors and strings connected to the ofuda device to suspend and pull it to add force vector to the hand of a participant. Figure 6 shows the ofuda device with tactile grip that installs six vibrators which provides tactile feedback sensation at the palm. The three dimensional position of the ofuda device is measured by the length of strings and its rotation by a spatial sensor. (Patriot, Polhemus)

In the course of the scenario of the VR space, the participant defended against the magic particles by a half-transparent shield with an amulet that is controlled by the ofuda device. The collision of particles caused force perturbation and impact and tactile stimulation on the device.

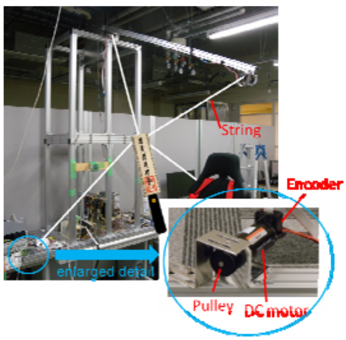


Fig. 5. Force feedback device

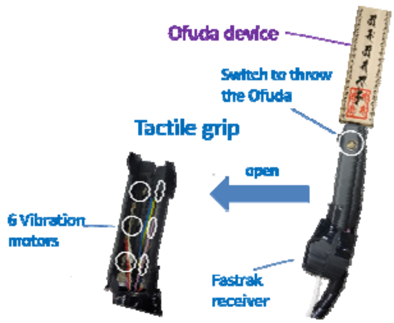


Fig. 6. Tactile feedback (ofuda) device

4.2 3D Motion Seat and 2D Feet Motion Generator System

Somesthetic (bodily) sensation system consists of a 3-dof motion seat (Fig. 7, 3D Seat) and 2-dof feet motion system (Fig. 8, Copernicus). This system produces vestibular stimulation that changes the participant's attitude (head position) in addition to the body posture change by the feet motion relative to the seat. The system generates both the changes in head position/orientation and in those of feet.

Fig. 7. 3D motion seat

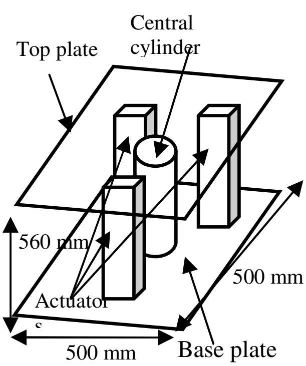


Fig. 8. Drive mechanism of the 3D motion seat

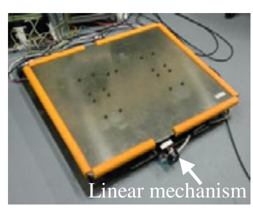


Fig. 9. 2D feet motion device (Copernicus)

The 3D motion seat shown in Fig. 7 consists of three electric linear actuators and a central cylinder (Fig. 8). The stroke of the actuator is 100 mm and its max velocity is 200 mm/s. The possible motions are up-and-down, pitch, and roll motions. These motions are combined to generate multiple dof motions of the seat. The 2D feet motion device shown in Fig. 9 installs two motors that drive the top plate 100 mm to the x-y directions in a horizontal plane where the participant puts the feet.

In the exhibition, the 3D seat and the feet device were driven simultaneously in accordance with the scenario to produce a whole body motion.

4.3 Tactile Feedback of the Seat

Tactile feedback was installed to work at the surface of the 3D motion seat (Fig. 10). The stimulator unit is a full-range speaker to create stimulations on the back and under the thigh of the seated participant to enhance the experience. Figure 10b shows two speakers installed in the bottom pad of the seat. Figure 10c shows six speakers in the back pad. These stimulators were driven so that the motion on the skin surface was properly created. In addition, the spatial sound of the particular scenario is mapped to enhance the effect of auditory information.

(a) (b) (c)

Fig. 10. a) Seat with tactile feedback, b) Stimulator units in the bottom pad, C) Six channel back vibrator unit

4.4 Wind and Scent Display

The stimulus of air flow to the face and upper body of a participant was controlled by eight wind fans (Figure 13). The fans were arranged in front of the participant with an appropriate interval to be able to change wind direction to his/her face. Eight to sixteen kinds of scent are able to be presented by the system. A scent of a scene is delivered by the compressed air that goes through one of the scent chambers which are filled with scent sources. The intensity of scent is determined flow volume by a valve control (Figure 15).

In the two exhibits, a scent display played a very important role since the questionnaire collected from the participants who had experienced with the demonstration rated that the scent was the most impressive among modalities.

Fig. 11. Wind device

Fig. 12. Chamber and compressor

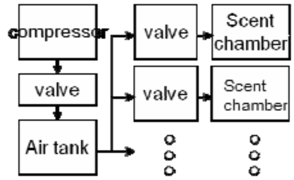


Fig. 13. System summary of the Scent display

4.5 Sound Presentation

A 7.1-ch sound was created around the participant with eight speakers. The sound of magic particles that the evil specter casted was presented at the position of particle obtained from VR environment software. In addition, some pieces of BGM music to the theme were composed and effectively enhanced the reality of the story.

4.6 Visual Presentation

The VR environment of this exhibition was rendered stereoscopically with a VR software, OmegaSpace (Solidray Co., Ltd.) and a stereo projector (Sight3D). The last scene which returns to the National Museum of Emerging Science was expressed by displaying realtime 3D image of the participant capturing with the stereo camera. The image was projected by a 3D projector and seen with shutter glasses.

5 Conclusion

The first stage implementation of Five Senses Theater (FiveStar) was investigated through the exhibition of Asiagraph. The rendered fantasy world was a good example to clarify how the multiple modality presentation was integrated effectively from the view point of ultrareality creation. We will continue to develop basic elements for multiple-modality presentation to improve the FiveStar performance. The interaction among modalities to enhance the whole effect of expression for communication and authoring content will be principal issue to pursue in this project hereafter.

Acknowledgments. This project has been supported by a grant for the ultra realistic communication technology by an innovative 3D image technique. We would like to thank National Institute of Information and Communications Technology, Japan.

References

1. Heilig, M.: Sensorama (1961), `http://en.wikipedia.org/wiki/-Sensorama`
2. Enami, K., Katsumoto, M., Nishimura, R.: Current techniques for Ultra-Reality and the future. The Journal of the Acoustical Society of Japan 64(5), 322–327 (2008) (in Japanese)

Synchronous vs. Asynchronous Control for Large Robot Teams

Huadong Wang[1], Andreas Kolling[1], Nathan Brooks[2],
Michael Lewis[1], and Katia Sycara[2]

[1] School of Information Sciences, University of Pittsburgh,
Pittsburgh, PA 15260 U.S.A.
[2] Robotics Institute, Carnegie Mellon University,
Pittsburgh, PA 15213 U.S.A.
huw16@pitt.edu, akolling@pitt.edu, nbb@andrew.cmu.edu,
ml@sis.pitt.edu, katia@cs.cmu.edu

Abstract. In this paper, we discuss and investigate the advantages of an asynchronous display, called image queue, for foraging tasks with emphasis on Urban Search and Rescue. The image queue approach mines video data to present the operator with a relevant and comprehensive view of the environment, which helps the user to identify targets of interest such as injured victims. This approach allows operators to search through a large amount of data gathered by autonomous robot teams, and fills the gap for comprehensive and scalable displays to obtain a network-centric perspective for UGVs. It is found that the image queue reduces errors and operator's workload comparing with the traditional synchronous display. Furthermore, it disentangles target detection from concurrent system operations and enables a call center approach to target detection. With such an approach, it could scale up to a larger multi-robot systems gathering huge amounts of data with multiple operators.

Keywords: Human-robot interaction, metrics, evaluation, multi-robot system, interface design, simulation.

1 Introduction

The task of interacting with multi-robot systems (MrS) especially large robot team presents unique challenges for the user interface designer, which is very different from the task of interacting with single or limited number of robots. Traditional graphical user interfaces and infrastructures have difficulties to interact with MrS. The core issue is one of scale: in a system of n robots, any task that has to be done to one robot must be done to the remaining n – 1 [1]. Interface for large robot team, should use an infrastructure that allows remote action management and centralized display of data. Many applications such as interplanetary construction, search and rescue in dangerous environments, or cooperating uninhabited aerial vehicles have been proposed for MrS. Controlling these robot teams has been a primary concern of many HRI researchers. These efforts have included theoretical and applied development of the Neglect Tolerance model and Fan-out model to characterize the control of independently operating robots [1, 2], predefined rules to coordinate

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 415–424, 2011.

cooperating robots as in Playbook™ [3] and Machinetta [4], and techniques for influencing teams obeying biologically inspired control laws [5, 6, 7]. While our efforts to increase span of control over unmanned vehicle (UV) teams appear to be making progress, the asymmetry between what we can command and what we can comprehend is growing. Automation can reduce excessive demands for human input, but throttling the information being collected and returned is fraught with danger. A human is frequently included in the loop of a MrS expressly to monitor and interpret video being gathered by UVs. This can be a difficult task for even a single camera [8] and begins exceeding operator capability before reaching ten cameras [9, 10]. With increasing autonomy of robot teams and plans for biologically inspired swarms of much greater size the problem of absorbing and benefiting from their product seems even more pressing than learning how to command them.

Foraging tasks, carried out with a large robot team, require an exploration that needs to be more than simply moving each robot to different locations in the environment. Acquiring a specific viewpoint of targets of interest, e.g. victims in a disaster scenario, is of greater concern and increasing the explored area is merely a means to this end. While a great deal of progress has been made for autonomous, exploration the identification of targets is still done by human operators who ensure that the area covered by robots has in fact been thoroughly searched for the desired targets. Without means to combine the data gathered by all robots the human operator is required to synchronously monitor their output, such as a video feed for each robot. This requirement and load on the human operator directly conflicts with other tasks, especially navigation which requires the camera to be pointed in the direction of travel in order to detect and avoid objects. The need to switch attention among robots will further increase the likelihood that a view containing a target will be missed. Earlier studies [11, 12] confirmed that search performance on these tasks is directly related to the frequency with which the operator shifts attention between robots, possibly due to targets missed in the video stream while servicing other robots.

The problem addressed in this paper is the design of an asynchronous, scalable, and comprehensive display, without requiring a 3d reconstruction, to enable operators to detect relevant targets in environments that are being explored by large teams of UGVs. We will present one particular design for such a display and test it in the context of USAR with large robot teams with some degree of autonomy and supervised by a single operator.

2 Asynchronous Imagery

2.1 Pilot Studies

An asynchronous display method can alleviate the concurrent load put on the human operator and disentangle the dependency of tasks that require the video feed. Furthermore, it can avoid attentive sampling among cameras by integrating multiple data streams into a comprehensive display. This in turn allows the addition of new data streams without increasing the complexity of the display itself. A first approach for an asynchronous display is explored in [13]. The method therein is motivated by asynchronous control techniques previously used in extraterrestrial NASA applications. These are faced with limited bandwidth and communication lags due to large interplanetary distances. Such lags make a direct operation and viewing of all outputs impossible. Instead, the robot team is instructed to gather information only at specific locations in the form of panorama images from an omnidirectional overhead camera. The operator then searches through all panorama images and determines the location of specific targets. The approach was tested in experiments which compared performance for operators controlling four robots in a team using streaming or asynchronous panorama displays. No significant improvement was found in terms of found victims, but the frequency of shifting focus between robots was correlated with performance for streaming video but not for asynchronous panoramas. As expected, the asynchronous display of information alleviates the need for excessive switching. It was conjectured that with larger robot teams the benefit of avoiding attention switching will increase and have an effect on performance in terms of victims. Further experiments in [14] scaling the team size to eight and twelve robots, however, found no further significant improvements. But this approach did not utilize all the available data from the video feeds that robots gather, so a huge amount of potentially useful information was discarded for the panorama condition. Furthermore, the operator needs to give the robots additional instructions where to sample panoramas.

In contrast to previous work our approach in this paper allows the use of autonomous exploration and we present an asynchronous display that mines all of the robot video feeds for relevant imagery. This imagery is then given to the operator for analysis. We coin this type of asynchronous display image queue and compare it to the traditional synchronous method of streaming live video from each robot (streaming video). In the next section we describe our test bed and followed by a detailed description of the image queue and a comparison with streaming video.

2.2 Image Queue Interface

The goal of the image queue interface is to best utilize the advantages of an asynchronous display and to maximize the amount of time human operators can spend on tasks which human's performs better than robots. Currently, for USAR, this is the case for tasks such as victim identification and navigating robots out of dangerous areas in which they got stuck. As the number of robots in a system increases with improved autonomy the demands on operators for these tasks increase as well. Hence, another requirement for the interface is to provide the potential for scaling to larger numbers of robots and operators. The proposed image queue interface implements the idea of asynchronous monitoring via a priority queue of images that allows operators to identify victims requiring neither synchronicity nor any contextual information not directly provided by the image queue.

The image queue interface (Fig. 1) focuses on two tasks: (1) viewing imagery and (2) localizing victims. It consists of a filmstrip viewer designed to present the operator with a filtered view of what has passed before the team's cameras. A filtered view is beneficial because the video taken contains a high proportion of redundant images from sequential frames and overlapping coverage by multiple robots.

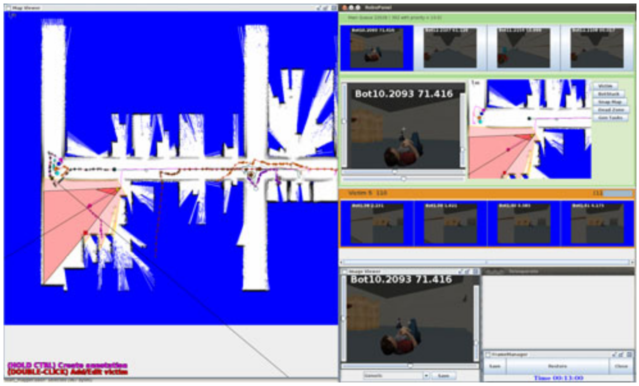


Fig. 1. GUI for the image queue condition

The filter attempts to reduce redundancy by only showing highly relevant images from the video stream. Relevance is scored by computing a utility for every image that determines its priority in the queue displayed in the filmstrip viewer. To achieve this we store every frame from all video streams in a database together with associated robot poses and laser scans taken at the time of capture. From this database we can retrieve any image and compute its utility. The computation of utility can be adapted to a particular application and for our experiment we computed utility via the area covered seen in an image. This visual coverage is computed by referencing the image in the map as seen in Fig. 2. Images with larger areas receive higher utility scores. Areas that have already been seen by other images in the filmstrip viewer do not count towards utility. In colloquial terms this kind of utility picks images that cover large areas with minimal overlap. Fig. 2 illustrates this concept of utility with a simple example.

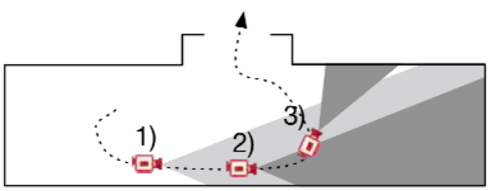


Fig. 2. An illustration of the utility of individual frames from a video stream. The frame taken at 1) has the largest visual coverage and highest utility while the frame at 2) has no utility since it is entirely overlapped by 1). Frame 3) has some utility since it provides coverage in an area not covered by 1).

By aggregating imagery with the highest utility scores at regular intervals the image queue allows the operator to peruse a relatively small number of prioritized images that show most of the new area explored by the robots. Notice that exploration can continue while operators view the image queue so long as robots are sufficiently autonomous (or controlled by other operators). Operators can either click through or scroll through a certain number of images in the queue. Once operators work through the first set of images the image queue marks the areas covered by these images as already seen and retrieves the next set of images with high utility. Tests of this system show that after 15 minutes of exploration an operator can view 70% of the area covered by viewing the 10 highest utility frames and 90% within the first 100 frames.

3 Methods

3.1 USARSim and MrCS

The experiment reported in this paper was conducted using the USARSim robotic simulation with 12 simulated Pioneer P3-AT robots performing Urban Search and Rescue (USAR) foraging tasks. USARSim is a high-fidelity simulation of urban search and rescue (USAR) robots and environments developed as a research tool for the study of human-robot interaction (HRI) and multi-robot coordination. USARSim supports HRI by accurately rendering user interface elements (particularly camera video), accurately representing robot automation and behavior, and accurately representing the remote environment that links the operator's awareness with the robot's behaviors. USARSim also serves as the basis for the Virtual Robots Competition of the RoboCup Rescue League.

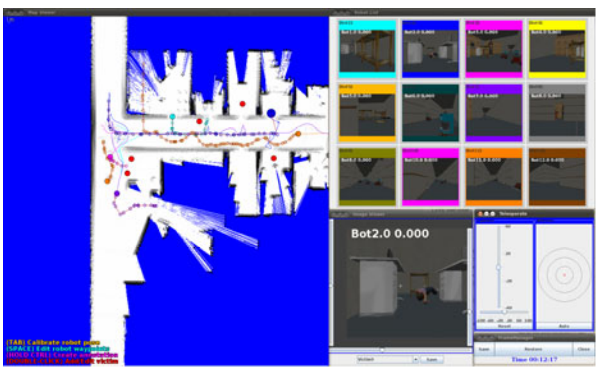


Fig. 3. GUI for the streaming video condition

MrCS (Multi-robot Control System), a multi-robot communications and control infrastructure with accompanying user interface, developed for experiments in multirobot control and RoboCup competition [15] was used in this experiment. MrCS provides facilities for starting and controlling robots in the simulation, displaying multiple camera, and laser output, as well as maps, and supporting inter-robot communication through Machinetta, a distributed multi-agent coordination infrastructure. Fig. 3 shows the elements of the conventional GUI for the streaming video condition. The operator selects the robot to be controlled from the colored thumbnails with live videos at the top right of the screen. The current locations and paths of the robots are shown on the Map Viewer (bottom left). Under manual control, robots are tasked by assigning waypoints on a heading-up map on the Map Viewer or through the teleoperation widget (lower right).

An autonomy path planner was used in the current experiment to drive the robots unlike the panorama study [13] in which paths were generated manually by participants with specified panorama locations. As the previous study [14, 17], operators appeared to have little difficulty in following these algorithmically generated paths and identified approximately the same numbers of victims (per operator) as in following human generated paths. A new Segment Voronoi Diagram (SVD) path planner replaced the random tree planner used in earlier studies. The new planner generates paths that maintain a safe distance to nearby obstacles. Such paths are generally longer, smoother and more human-like.

3.2 Experimental Design

The experiment followed a two condition repeated measures designs comparing the conventional MrCS displays (streaming video) with MrCS augmented by the experimental image queue display counterbalancing conditions. Automated path planning to improve search performance and autonomous exploration was used in both conditions. The operators performed a supervisory control task in which the robots navigated autonomously with the operator allowed to override by directing them through new waypoints. When necessary, participants were able to teleoperate the in-focus robot to extricate it when it became stuck.

32 paid participants were recruited from the University of Pittsburgh community balanced among conditions for gender. None had prior experience with robot control although most were frequent computer users.

After providing demographic data and completing a perspective taking test participants read standard instructions on how to control robots via MrCS. In the following training session, participants practiced control operations for both streaming video and image queue condition for 10 min each. Participants were encouraged to find and mark at least one victim in the training environment under the guidance of the experimenter. After the training session, participants began the two 15 minute real task sessions in which they performed the search task controlling 12 robots in teams using either the streaming video or image queue display with a counterbalanced design. At the conclusion of each real task session, participants were asked to complete the NASA-TLX workload survey [16].

4 Result

Data were analyzed using a repeated measures ANOVA comparing streaming video with the image queue condition. Overall, in both conditions participants were successful in searching through the environment. On average participants in the streaming video condition found 9.10 victims while those in the image queue condition found 8.51 (Fig. 4) without a significant difference between conditions ($F_{1,28} = .733$, $p = .387$). In addition, as Fig. 4 shows, area explored by the 12 robots was also showed a none significant advantage for the streaming video condition, ($F_{1,28} = 2.147$, $p = .154$) .

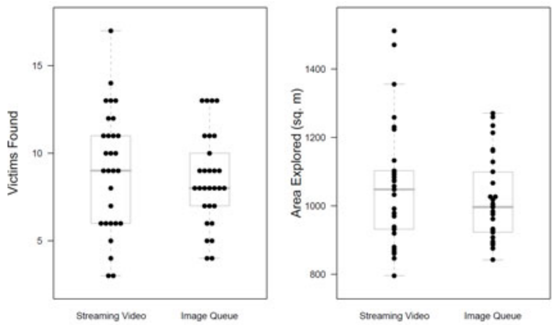


Fig. 4. Victims found and area explored

Victim marking errors have been classified into two types, false positive and false negative. A mark without any victim or multiple marks for one victim were was counted as a false positive. The number of false positive shows a significant advantage for image queue condition than the streaming video condition ($F_{1,28}$ = 13.032, p = .001). On the other hand, Victims that were missed, but present in the video feed, and not marked were calculated as a false negative. The image queue shows a similar significant improvement for the missing victim comparing with the streaming video condition($F_{1,28}$ = 5.526, p = .026), with the average number of missing victims in the image queue condition dropping to 7.48 while participants in the streaming video condition missing 9.34 victims respectively (Fig. 5).

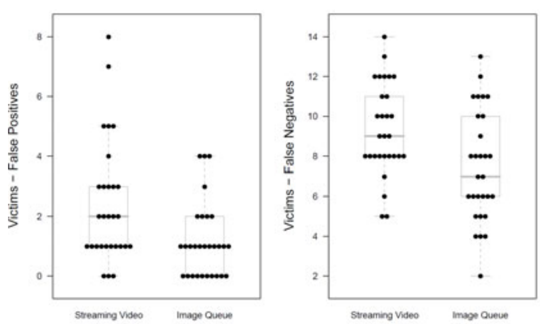


Fig. 5. Marking errors of victims

To more closely measure the subjects behavior linked to their monitoring and operation, teleoperation times (Fig. 6) were analyzed. The repeated measures ANOVA shows a significant difference for the count of teleoperation times between the streaming video and image queue condition, with participants in the streaming video condition, teleoperate average 21.24 times, which is significantly more that the participants in the streaming video condition, teleoperate 4.97 times respectively ($F_{1,28}$ = 150.719, p < .001).

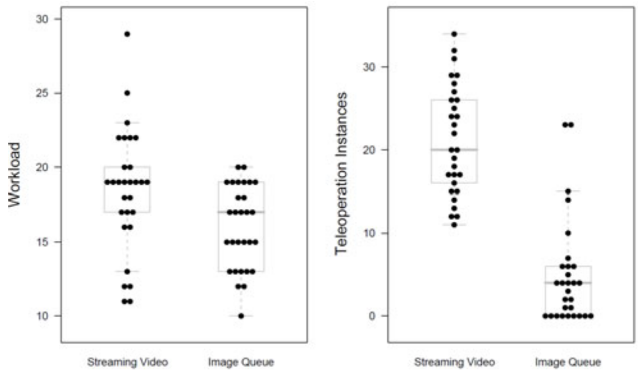


Fig. 6. Teleoperation and workload

The full scale NASA-TLX workload measure also revealed a significant difference, which is unlike the earlier studies [28] where no advantage was found for panorama pictures GUI. A significant advantage in workload ($F_{1,28}$ = 7.347, p = .001) was observed favoring the image queue condition (Fig. 6). Examining individual dimension of workload comparing streaming video and image queue we found significant differences for temporal demand ($F_{1,28}$ = 6.503, p = .016) and effort ($F_{1,28}$ = 4.576, p = .040).

5 Discussion and Future Works

The purpose of this experiment was to examine the asynchronous image queue interaction method to control a large human robot teams. Our experiment result shows that under image queue conditions, which allowing interruption, relevance image retrieval, reviewable, location based image queue interface leads to somewhat equal search performance with lower errors and lower workload.

In the image queue condition humans were able to view the image queue and control 12 robots each, to mark the equal number of victim and to cover the equal area as the streaming video condition participants. This result supports the feasibility of our image queue interaction mode as a substitute for human operators (performance is no worse). The analysis of victim marking errors shows that in the streaming video condition participants marked 89.3% more victims at the wrong location and missed 23.1% more victims than image queue condition. This gain for image queue is particularly significant because avoiding missed targets is crucial to many foraging tasks, thoroughness and correctness is more important than other performance.

Similar improvements in reported workload suggest that substantial cognitive resources were required for streaming video condition. Based on the analysis result of teleoperation instance, participants in streaming video condition are more likely to intervene the robot for path planning and problem solving. Participants in streaming video condition have teleoperate 4.8 times more than those in image queue condition during the whole task.

Participants in the streaming video condition were confronted with a bank of videos (Fig. 3) much like a security guard monitoring too many surveillance cameras. Informal observation of participants suggested that due to the frequently distraction of robot operation requirement, victim running into the camera and task switch in between, operator puts great effort for task allocation and feels intensive time pressure. The significant difference of individual dimension of workload, temporal demand and effort, supports this observations.

Furthermore, while we undertook this study to determine whether asynchronous video might prove beneficial to larger teams we found performance to be essentially equivalent to the use of streaming video at big team sizes with lower errors and workload. When people doing search tasks under stressed or heavily loaded conditions spontaneous self-synchronization is unlikely to arise and hence full shared information. Instead, image queue provided a benefit way of task division between operators, in which exploration and perceptual search (identifying targets) task could be neatly deconstructed to different operators. Framing the problem this way leads to the design conclusion that operators should be issuing task-centric commands. To realize this kind of control architecture we will propose a call center approach in which some operators address independent control needs for monitoring and exploration of UVs, whereas other operators address independent location based images in queue for victim marking and other perceived tasks.

Once we are facing the problem of managing 12 or more robot teams or in other words, the large scale systems, the call center design of the control architecture with image queue embedded will be proposed to solve these problems. Then information sharing problem with people's losing of situation awareness will be one of the priorities of our concern. Just imaging three operators controlling more than 30 UVs with some UV originated requests such as verifying/marking targets or alerting operator on low fuel and other UV needs such as veering off path or becoming bogged down by terrain (trees in a forest for example) or similar that operators' must monitor to detect; as a result the image queue interaction mode may still be a good choice for people to share information or recall for their situation awareness.

According to these concerns, we want to explore the effect the combination of these two modes of control in large number of robots with multiple operators, in which under a call-center structures exploration task will be operated in the streaming video mode and perceptual search task will be performed in the image queue mode, and it could be useful in preparing to study larger scale systems in the future.

Acknowledgments. This research has been sponsored in part by AFOSR FA955008-10356 and ONR Grant N0001409-10680.

References

1. Olsen, D., Wood, S.: Fan-out: Measuring Human Control of Multiple Robots. In: Human Factors in Computing Systems (CHI 2004). ACM Press, Vienna (2004)
2. Crandall, J., Goodrich, M., Olsen, D., Nielsen, C.: Validating human-robot interaction schemes in multitasking environments. IEEE Transactions on Systems, Man, and Cybernetics, Part A 35(4), 438–449 (2005)

3. Miller, C., Parasuraman, R.: Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. Human Factors 49(1), 57–75 (2007)
4. Scerri, P., Liao, E., Lai, L., Sycara, K., Xu, Y., Lewis, M.: Coordinating very large groups of wide area search munitions. Theory and Algorithms for Cooperative Systems
5. Kira, Z., Potter, M.: Exerting Human Control Over Decentralized Robot Swarms. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2010 (2010)
6. McLurkin, J., Smith, J., Frankel, J., Sotkowitz, D., Blau, D., Schmidt, B.: Speaking Swarmish: Human-Robot Interface Design for Large Swarms of Autonomous Mobile Robots. In: AAAI Spring Symposium (2006)
7. Ding, X., Powers, M., Egerstedt, M., Young, S., Balch, T.: Executive Decision Support: Single-Agent Control of Multiple UAVs. IEEE Robotics & Automation Magazine 16(2), 73–81 (2009)
8. Cooke, N., Pringle, H., Pedersen, H., Connor, O. (eds.): Human Factors of Remotely Operated Vehicles. Elsevier, Amsterdam (2006)
9. Humphrey, C., Henk, C., Sewell, G., Williams, B., Adams, J.: Assessing the Scalability of a Multiple Robot Interface. In: 2nd ACM/IEEE International Conference on Human-Robotic Interaction (2007)
10. Lewis, M., Wang, H., Chien, S., Velagapudi, P., Scerri, P., Sycara, K.: Choosing autonomy modes for multirobot search. Human Factors 52(2), 225–233 (2010)
11. Pepper, C., Balakirsky, S., Scrapper, C.: Robot Simulation Physics Validation. In: PerMIS 2007 (2007)
12. Taylor, B., Balakirsky, S., Messina, E., Quinn, R.: Design and Validation of a Whegs Robot in USARSim. In: PerMIS 2007 (2007)
13. Velagapudi, P., Wang, J., Wang, H., Scerri, P., Lewis, M., Sycara, K.: Synchronous vs. Asynchronous Video in Multi-Robot Search. In: First International Conference on Advances in Computer-Human Interaction, ACHI 2008 (2008)
14. Velagapudi, P., Wang, H., Scerri, P., Lewis, M., Sycara, K.: Scaling effects for streaming video vs. static panorama in multirobot search. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS (2009)
15. Robocup Rescue VR, `http://www.robocuprescue.org/wiki/index.php?title=VRCompetitions#Singapore_2010`
16. Hart, S., Staveland, L.: Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 139–183. Elsevier, Amsterdam (1998)
17. Lewis, M., Wang, H., Chien, S., Velagapudi, P., Scerri, P., Sycara, K.: Choosing autonomy modes for multirobot search. Human Factors 52(2), 225–233 (2010)

Acceleration of Massive Particle Data Visualization Based on GPU

Hyun-Rok Yang, Kyung-Kyu Kang, and Dongho Kim

Department of Media
Soongsil University
511 Sangdo-dong Dongjak-gu, South Korea
{penril87,rcrookie}@magiclab.kr, cg@su.ac.kr

Abstract. In case of using ray tracing for rendering of massive particles, it usually takes more time than mesh-based models, where the particle data used for the representation of fluid are generated by the fluid simulator. It also has a large amount of particle set with high density. In this paper, we apply two schemes to reduce the rendering time and solve the problems using the characteristics of the particles. We suggest that GPGPU can improve the efficiency of the operations through parallel processing and a modified Kd-tree spatial subdivision algorithm for speed-up.

Keywords: GPGPU, CUDA, Ray tracing, Kd-tree, octree.

1 Introduction

Particle data sets are used to represent fluid such as smoke, water, or flame. Massive particles are created by fluid simulators in order to express realistic natural phenomena [1]. However the particle data have a large amount of particles with high density.

The particle data usually amounts to millions to billions in film making industry. Due to this characteristic of particle data, rendering speed becomes slow using a ray tracing method due to the number of ray-object intersection tests required. Figure 1 shows the rendering result of the particle data and the density of particles is high in the box sections.

We can easily find that some particles are located with high density only in very small number of areas. This seriously imbalanced distribution causes some problems when we construct spatial subdivision structures for ray tracing. Some nodes of the structure have massive particles than others, which can increase the rendering time a lot.

This paper takes two approaches to solve this problem and reduce the rendering time for particle data ray tracing. First, we employ a ray tracing technique based on GPU for the rendering speed-up. Recently, GPUs have hundreds of streaming processing units, so they are suitable for the acceleration of ray tracing. Second, we can take advantage of spatial subdivision structures such as Kd-tree, octree, or uniform grids. We use a modified Kd-Tree, where octree data strcuture is additionally used in order to subdivide and manage particle data for some leaf nodes of Kd-tree. Although it may increase the traverse time on the tree, it will decrease the number of ray-particle intersection tests, which finally reduces the rendering time.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 425–431, 2011.

Fig. 1. Dense particle data in boxes (gray rectangles)

2 Related Works

In this section, we review the previous parallel computing techniques based on GPU.

2.1 Parallel Computing Architectures Using GPU

In GPU computation, data caching and flow control are not important when compared to the CPU, because GPU is designed for parallel computation and high efficiency computation. In addition, GPU has more transistors than CPU for data processing. Many techniques have been developed using these characteristics of GPU efficiently [2].

CUDA(Compute Unified Device Architecture) is a programming language developed by nVIDIA. CUDA is a C-like language allowing high-efficiency operations based in GPU, as an nVIDIA's parallel computing framework.

With millions of CUDA-enabled GPUs sold, software developers, scientists, and researchers are trying to utilize CUDA in many areas, including image and video processing, computational biology and chemistry, fluid dynamics simulation, CT image reconstruction, seismic analysis, ray tracing, and many more [3]. There are also similar GPGPU architectures such as OpenCL by Khronos Group and Direct Compute by Microsoft.

2.2 Parallelization Researches Based on GPU

Many researches use CUDA for GPU programming, because nVIDIA provides good technical support and the CUDA library is easier to use than other technologies such as OpenCL and Direct Compute. It is main reason that we used CUDA in this research.

Recently, Shubhabrata and his colleagues researched on the parallelization of scan algorithm [4]. The work consists of three parts for the parallelization using CUDA, - 1) Define work per each thread, 2) Store the results from the threads of each block in parallel, 3) Store the results from the blocks of each grid in parallel. We adopted this scheme, because we had to utilize the concepts of CUDA parallelization such as threads, blocks and grids.

Nan Zhang and his colleagues also implemented GPU-based parallel image processing algorithms following the steps mentioned above to accelerate the Sobel edge detection algorithms [5].

3 GPU-Based Acceleration Techniques for Ray Tracing

In this paper, we present two methods to accelerate ray tracing and solve the issues caused by the characteristics of particles' unbalanced distribution in a scene. First, we parallelize the existing ray tracing algorithm to maximize of computational efficiency using GPU. Second, we propose a modified Kd-tree algorithm with octree data structure attached on the leaf nodes, to solve the particle imbalance problem. In this section, we explain how the algorithm works in detail.

3.1 Parallelization of Existing Algorithms

We developed an off-line renderer which renders massive particle data using parallel ray tracing based on GPU. Recursive functions are not used at all, because they are not suitable in GPU computation and may cause GPU memory leakage. In our renderer, ray shooting operations are performed for all pixels according to the resolution, and they are parallelized following CUDA's parallelization schemes (thread, block, grid) mentioned in the previous existing researches [4, 5].

A thread corresponds to a ray. If we assume that we shoot only one ray per each pixel, a thread can cover a single pixel. In our experimental results, a recommended setup is that a block consists of 64 threads, which means that one block covers 8x8 pixels. For the grid, a good setup was that a grid covers the whole pixels the screen with the required number of blocks. This setup could produce the best rendering speed in our experimental environment.

Each thread performs a ray-object intersection test. Since we represent a particle as a sphere, we implemented a ray-sphere intersection test algorithm based on GPU.

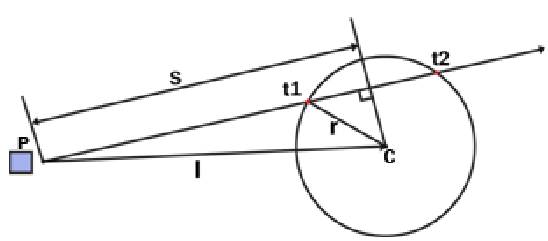


Fig. 2. Ray-sphere intersection test [6]

Figure 2 shows how the ray-sphere intersection test works. The intersection test algorithm can detect whether the sphere intersects with the ray fired from the pixel P and where the intersection occurs on the sphere.

A sphere is represented by the center C and radius r in this algorithm. Vector I is the vector from the origin to the center of the sphere. After obtaining this vector, we can determine whether the ray was off from the sphere or not, according to S, the dot product I and the ray direction.

If there is an intersection, we can compute where the actual intersection occurs. Between the two intersection points, we use t1, which is the closer intersection from the origin.

3.2 The Modified Kd-tree Data Structure

Kd-tree is one of the acceleration algorithms for ray tracing [7]. However, if massive particle data are managed only by Kd-tree, as shown in red box nodes in Figure 3, unbalanced population cases may occur on some nodes.

In this paper, in order to solve this problem, if the number of particle data in a Kd-tree node exceeds the threshold, we subdivide the nodes further to manage the particle population. Our method, first, builds Kd-tree structure with all particle data, then performs further subdivision in the nodes with dense population.

Construction of the Hybrid Tree Structure. We first build the Kd-tree structure with median split strategy. The strategy subdivides the space by the median plane along the longest axis of the aligned bounding box for the given particle data.

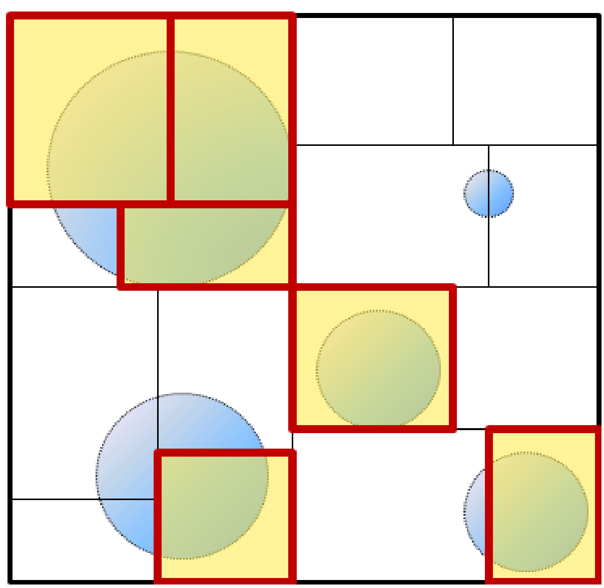


Fig. 3. Particle density imbalance problem at some nodes of the Kd-tree
* Spheres represent particle populations.

After the construction of the Kd-tree, additional subtrees are added to some leaf nodes of the Kd-tree. The subtrees, are octrees which divides the node into 8 subspaces using three axis planes in x, y, and z. A subtree is added only in the node where the number of particles in it exceeds the given threshold. The subtree data structure is stored in one dimensional array, which simplifies the traversal using GPU.

Tree Traversal. The tree traversal in our work is simply a variation of conventional Kd-tree traversal. When the Kd-tree traversal reaches a leaf node, it is checked if the node contains a subtree or not. If the node has a subtree, the subtree is traversed using the bounding boxes of the tree's nodes.

When the ray intersects with the bounding box, the ray sphere intersection tests are performed. In order to accelerate the test, we use the algorithm suggested by Smits [8]. This method checks the intersection between the ray and three slabs instead of six faces of the bounding box.

4 Experimental Result and Analysis

For the performance analysis of our acceleration algorithm based on GPU, a PC was used equipped with Intel i7 CPU, 12GB RAM, and nVIDIA geForce GTX480 graphic card with 1.5GB video memory. The rendering resolution is 1920 pixels by 1024 pixels, and 4 light sources are used. For performance analysis, we used various scenes with various depths of Kd-tree and subtree (octree). A memory usage was also measured for the test scenes.

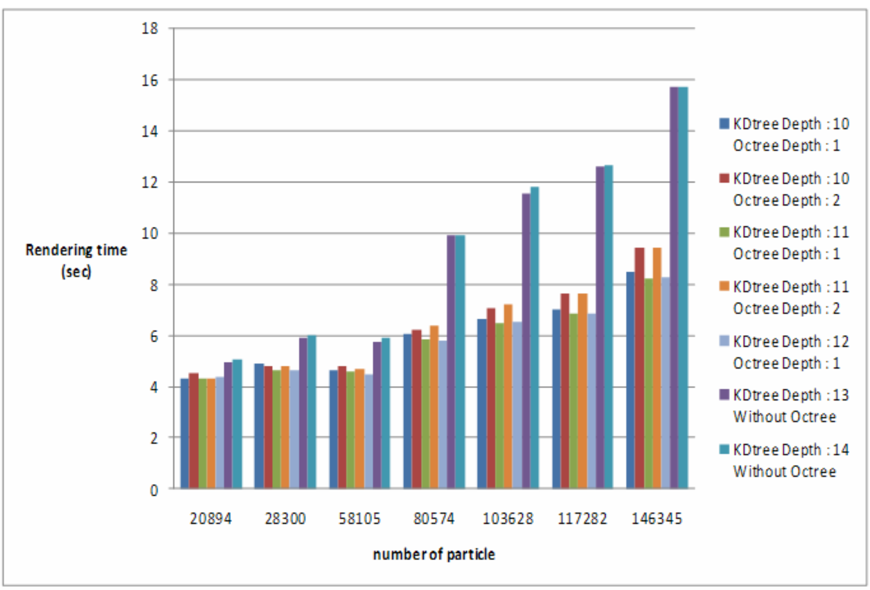


Fig. 4. Performance comparison

Figure 4 shows the comparison of rendering time according to the depths of Kd-tree and octree. In this result, rendering is faster when Kd-tree and octree are used together than only Kd-tree is used.

The speed-up comes from the fact that octree can subdivide the space more precisely than Kd-tree, thus the numbers of particles in the nodes are reduced. But, we cannot use octree only, because so many nodes should be traversed, which makes the rendering time even slower. Moreover, more memory is required in octree subdivision.

In this research, octree subdivision is performed only for some leaf nodes in Kd-tree, and the depth of octree is limited only up to 3. As a result, the advantage of using octree could be taken based on the modified Kd-tree without using much memory.

Memory usage is important because the amount of memory is restricted for GPU in most cases. Figure 5 shows the memory consumption according to the octree depths. When the depth of Kd-tree is 10 and the octree depth is 1, the memory usage is similar to the case that only Kd-tree is used with depth of 13.

This analysis shows that the speed-up can be obtained when we decrease the depth of Kd-tree and subdivide the leaf nodes further using octree, without increasing the required memory.

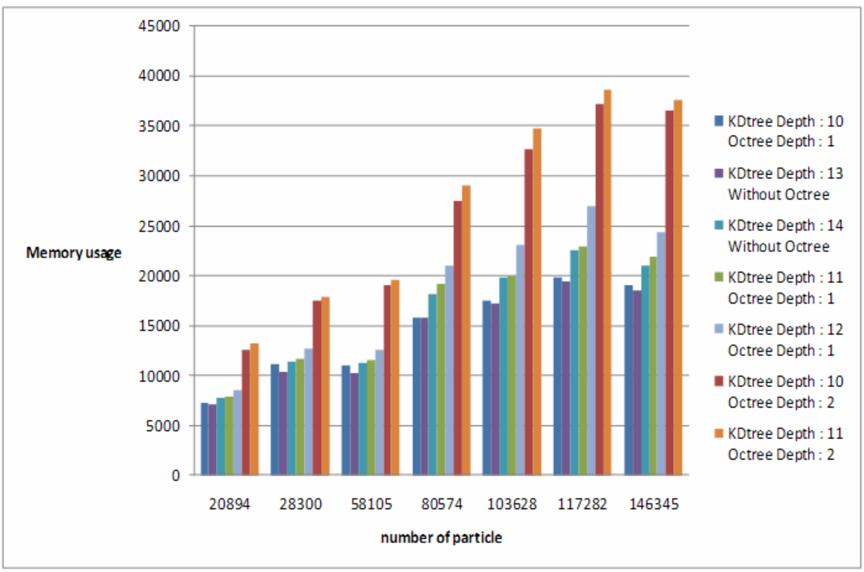


Fig. 5. Comparison of memory usage

5 Conclusion

We showed that rendering time can be reduced using both Kd-tree and further octree subdivision rather than using Kd-tree only. Moreover, memory usage is more effective than using Kd-tree only since we can balance the population of particle data. In our experimental results, a large depth of the subtree is not efficient because it may only increase the traversal time.

Since our approach manages the space subdivision by combining Kd-tree and octree subdivision, it was able to improve the rendering speed by adjusting the depth of Kd-tree and subtree based on the numbers of particles in the nodes. While our current approach is using one GPU, the utilization of multi-GPUs may be a good choice for further speed-up in the future.

Acknowledgments. This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2009.

References

1. Shin, S.H., Yoon, J.C., Lee, J., Kim, C.H.: Quadrangular Splatting based on SPH. Korea Computer Graphics Society 28, 1691–1722 (2008)
2. nVIDIA.: nVIDIA CUDA C Programming Guide. nVIDIA Corporation (2010)
3. What is CUDA? http://www.nvidia.com/object/what_is_cuda_new.html
4. Sengupta, S., Harris, M., Garland, M.: Efficient Parallel Scan Algorithms for GPUs. nVIDIA Technical Report (2008)
5. Zhang, N., Chen, Y.S., Wang, J.L.: Image Parallel Processing Based on GPU. In: Advanced Computer Control (ICACC), pp.367–370 (2010)
6. Tomas, A.M.: Real Time Rendering, 2nd edn. AK Peters (2003)
7. Bentley, J.L.: Multidimensional binary search trees used for associative searching. Communications of the ACM 18, 509–517 (1975)
8. Smits, B.: Efficiency issues for ray tracing. In: ACM SIGGRAPH 2005 Courses (2005)

Author Index